





Db 929 QOETASRELVKEPARAGDRQPEWLEEQQGRFQSTQAALQAM--EREAQMGNELERLRA 986  
QY 197 ATAAQGGYEIPARLRLHNLVIOYAAQGRYEVAVPLCKQALEDLERTSGRHGPDVATML 256  
Db 987 ALMESQGO-----QOEEGQOE-----REVARLTQERGRAQADLAEK 1024  
QY 257 NILA-LVYRDQNKYE-----AAHLLNDALSIRESTLGPDPHPAATLNNLAVLYG-K 307  
Db 1025 AARAELEMLQNALNEQORVEFATLQEAHALHTEKE---GKD-----QELAKLRGLE 1073  
QY 308 RGYKEAEPLCQRALEIREKV-----LGTNHPDVAK-----QLNNLA 344  
Db 1074 AAQIKELEELRQTVKQLKEQLAKKEHAGSGAQSEAAAGRTPTGPKLEALRAEVSKLE 1133  
QY 345 LLCQNGKY-EAVERYQ-----RALAIYEQ-----LQPDNPNVARTKNLA 386  
Db 1134 QQCQKQEQADSLERSLEAERASRAERDSALETLQGQLEKAQELGHSQSALASAQRELA 1193  
QY 387 SCYLKQGYAEATLYK-----EILTRAHVQEFGSVDDHK 422  
Db 1194 AFRTKVQDHSKAEDENKQAVARGQAEAKNLSLSEEEVSILNRQVLEKEGESKELKR 1253  
QY 423 PIWMHAEREEMSKSRHHEGGTPYAEYGGWYKACKVSSPTVNTTLRNLGALYRRQKLEA 482  
Db 1254 LVMAESEKSQKLEES-----CACCRQROPATVPELQNALLCGRRCRASG 1298  
QY 483 AETLEECALSRROGWDPISQTKVAELLGESDGRRTSQEGPGDSVKFEGGEDASVAVWS 542  
Db 1299 REA-EKQVASENLRQELTSQAERAEELG-----QELKAWQEFQKEQALSTLQLE 1349  
QY 543 GDGSGTL-----QSGSLGKJR-DVLRSSSEL--L 569  
Db 1350 HTSTQALVSELLPAKHLCCQLQAEQAAAEKRHRELEQSKQAAGGLRAELLRAQRELDEL 1409  
QY 570 VRKLOGTEPRPSSSNMKRAASLNLYLNQPSAAPLQVSRGLSA 610  
Db 1410 IPLRQKVAEQERTAQOLRAEKASYAEQLSM--LKKAHGLLA 1448

RESULT 4  
W00024  
ID W00024 standard; Protein; 1972 AA.  
AC W00024;  
DT 25-MAR-1997 (first entry)  
DE Smooth muscle myosin heavy chain SM1 isoform protein.  
KW Smooth muscle; myosin heavy chain; SM1 isoform; rabbit; arteriosclerosis;  
KW gene therapy; mouse; SM2 isoform; retrovirus; adenovirus; restenosis;  
KW associated adenovirus; coronary artery catheterisation; sclerotic artery.  
OS Mus musculus.  
PN W09623069-A1.  
PD 01-AUG-1996.  
PF 25-JAN-1996;  
PR 25-JAN-1995; JP-010085.  
PA (OSAP ) OSAKA PREFECTURAL GOVERNMENT.  
PA (VESS-) VESSEL RES LAB CO LTD.  
PI Arakawa E, Hasegawa K, Ishiyama H, Matsuda Y, Oda S;  
PI Sugawara M, Takahashi K;  
DR WPI; 96-362693/36.  
DR N-PSDB; T34291.  
PT DNA encoding smooth muscle myosin heavy chain SM1 isoform - used in  
PT a vector for gene therapy for reduction of re-stenosis following.  
PT coronary artery catheterisation  
PS Claim 5; Page 14-27; 42pp; Japanese.  
CC This sequence represents the smooth muscle myosin heavy chain SM1 isoform  
CC protein. The SM1 coding sequence was isolated from a mouse embryo cDNA  
CC library using probes based on the rabbit smooth muscle myosin heavy chain  
CC SM2 isoform. The isolated coding sequence is included in the gene  
CC therapy vector of the invention. The vector of the invention consists of  
CC the coding sequence inserted into a retrovirus, adenovirus, associated  
CC adenovirus or animal-expression plasmid vector (such as pCXN2 or  
CC PAGE208). The vector can be used in the gene therapy treatment of  
CC arteriosclerosis, particularly for the reduction of restenosis occurring

CC following coronary artery catheterisation for widening of sclerotic  
CC arteries. 1972 AA;  
SQ Sequence  
  
Query Match 5.3%; Score 166; DB 1; Length 1972;  
Best Local Similarity 20.7%; Pred. No. 3.5e-05;  
Matches 140; Conservative 120; Mismatches 278; Indels 138; Gaps 25;  
  
QY 5 VLQORDEPAGHRLSQEETIGSTRVLSQGLEALRSEHQAVLOSLSQTIECLOQGGHE-EG 63  
Db 1242 VLQQAQVEVEHKKKLEV-----QLQDLQSKSDGERARAELSDKVHKLQNEVESVTGM 1295  
QY 64 VHE---KARQLRRSMENI-----ELGLSEAQVMLALASHLSTVESEKQLRAQVRL 112  
Db 1296 LNEAGKAIKAKDVASLGSQQLQDTQELLQEEFRQKLNSTKLRLQLEDERNSLQDLDEE 1355  
QY 113 CQENQ-----WLRDELACTQRLQRSEQAQVAALEEEKKHL--EFLGQLRQYDED--GH 161  
Db 1356 MEAQNLERHVTNLNIQSDSKKKLQDFASTIEVMEEGKKRLQKEMEGLSQQYEEKAAAY 1415  
QY 162 TSEEKEGDATEKSLDDLPFN-EEEDPSNGLSRGQGATAAQGGYEIPARLRLHNLVIQ 220  
Db 1416 DKLEKTKNRLQOELDDLVDLDNQRLVSNLEKKQ-----KKFDQLLAEKKNISSK 1466  
QY 221 YA-----AQGRYEVAVPLCK-----QALDERTSGRHGPDVATMLNIALVYRD 265  
Db 1467 YADERDRAEAEAREKETKALSARALEEALAEKEELERTNKMMLKAEMEDLVSSKDDVGKN 1526  
QY 266 QNKYKEAAHLLNDALSIRESTLGPDPHPAATLN-----NLAVLYGKRGKYKEAEPLC 318  
Db 1527 VHELEKSKRALETQMEEMKTQLESEDDVQATEDAKLRLEVNMQALKGQ-----F 1576  
QY 319 QRALEIREKVLGTNHPDVAKQLNNLALICQNGKYEAVERYYQRALAIYEGQLGPDNPV 378  
Db 1577 ERDLQARDE-----QNEEK---RRQLQRQLHEYEYETELEDERKQR 1612  
QY 379 A-----RTKNNLASCYLK-----QGYAEAEATLYKEILTRAHVQEFGSVDDHKKPIWM 426  
Db 1613 ALAAAANKKLEGDLKLELOADSAIKGREAIAKQLRK---LQAOMKDFQRELDAR---- 1665  
QY 427 HABEREEMSKSRHHEGTPYAEYGGWYKACKVSSPTVNTTLRNLGAL--YRRQKLEAAE 484  
Db 1666 -ASRDEIFATSKENE-----KKAKSLEADLMQLQEDLAAERARKQADLEKEE 1712  
QY 485 TLEECAL---RSRQGTDPISQTKVAELGESDGRRTSQEGPGDSVKFEGGEDASVAVE 540  
Db 1713 LAELASSLSGRNTLQDEKRRLEARIAQLEEELEEEQGNMEAMSDRVRKATLQAEQLSNE 1772  
QY 541 WSGDGSGLTQRSGSLKIRDVLRRSSELLVRKIQGTGTEPRPSSSNMKRAASLN-YLNQPSA 599  
Db 1773 LATER-SQAQKNES---ARQQLERQNKELRSKIQVEGAVKAKLKSIVAALEAKIAQLEE 1828  
QY 600 APLQVSRGLSASTMDL 615  
Db 1829 QVEQEAREKQAATKSL 1844

RESULT 5  
W46499  
ID W46499 standard; Protein; 981 AA.  
AC W46499;  
DT 19-MAY-1998 (first entry)  
DE Amino acid sequence of the spindly (SPY) protein of Arabidopsis.  
KW Gibberellin signal transduction; spindly phenotype; SPY gene; rescue;  
KW spy mutant gene; gibberellin overdose syndrome; paclobutrazol; spy-4 DNA;  
KW modulation; plant development; plant height; fruit growth;  
KW flower development; leaf size.  
OS Arabidopsis thaliana.  
FH Key Location/Qualifiers  
FT Region 1..76  
FT /note= "N-terminus"  
FT Domain 77..430







QY 134 EQAVAQLEEEKHL-EFLGQLRQYDEGHTSEKEGDATKDSLDDLPNEEEDPSNGL- 191  
Db 800 RRAADALEEQRCISELKAETRSVLEQ-HKRRKE-----LEERAGRKGLE 845  
QY 192 SRQGATAAQGGYEIPAR-----LRTLH-----NLVIQYAA-QGRYE----- 228  
Db 846 ARLOQLGEAHQAETEVLRRLAEAMAAQHTAESECEQLVKEVAAWRVEDSQEEAQYG 905  
QY 229 -----VAVPLCKQALEDL-----ERTSG-----RGHPDVATM-LNILALVYRD 265  
Db 906 AMFQELMTLKEEKEKARQELQEAKEKVAGIESHSELQISRQQNELAEHLANLALQV 965  
QY 266 QNKYKEAAHLLNDALSIRESTLGPDPHAPAAATLNNLAVLYGKRGYKEAEPLCQRALEIR 325  
Db 966 QEKEVRAQKLADDLSTLQEK-----MAATSKVARLETILVRKAGEQETASRELVE 1017  
QY 326 EKVLTGNHPDVAKQLNNLALLCQNGKYEAVERYQYR-----ALAIYEGQ-----L 371  
Db 1018 PARAGDROPEWLEEQG-RQFCSTQAAALQAMEREAEQMGNELERLRAALMESQGOQEEER 1076  
QY 372 GPDNPNVAR--TKNNLASCYLKQGYAEAE-----TLYKEILTRAHVQE 413  
Db 1077 GQEREVARLTQERGRAQADLALEKAARAELEMLRLQNALNEQVFEFATLQEAHALTEK 1136  
QY 414 FGSVDDHDKPIWMHAEEREEMSKSR-----HHEGGTPYAEYGGWKACK 457  
Db 1137 EGKQELAKLRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHAGSGAQSEAAAG---RTE 1193  
QY 458 VSSPTVNTTLRNLGALYRR-QGKLEAAETLEEC--ALRSRRQGTDPISQT-----KVAE 508  
Db 1194 PTGPKLEALRAEVSKELEQCCQKQEQADSLERSLEAERASRAERDSALETLQGLEEKAQ 1253  
QY 509 LLGESDGRRTSQEGPGDSVKFEGGEDASVAVEWSGD---GSGTLQRSGLSKIRDVLRRS 565  
Db 1254 ELGHSQSALASAQRELAAFRTKVQDHSKAEDENKAEDEWKAQVARGRQEAERKNSL-----ISSLE 1308  
QY 566 SELLVRKLOGTEPRPSSSNMKNR 587  
Db 1309 EVSILNRQVLEKEGESKELKR 1330

RESULT 7  
W21731  
ID. W21731 standard; Protein; 2272 AA.  
AC W21731;  
DT 01-OCT-1997 (first entry)  
DE GAL4/HA/NuMA fusion protein.  
KW NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;  
KW cell division; proliferation; antibody; Ab; detection;  
KW malignant cell growth.  
OS Homo sapiens.  
FH Key  
FT domain  
FT 1. .147  
FT /label= GAL4\_DNA\_binding\_domain  
FT 148. .174  
FT /label= Hemagglutinin epitope  
FT 175. .2272  
FT /label= Residues\_18-2116\_of\_NuMA  
FT 365. .1864  
FT /label= Coiled\_coil\_region  
PN WO9640917-A1.  
PD 19-DEC-1996.  
PE 07-JUN-1996; U09504.  
PR 07-JUN-1995; US-478408.  
PA (UYVA ) UNIV YALE.  
PI McPherson SMG, Snyder MP;  
DR WPI; 97-077270/07.  
DR N-PSDB; T77782.  
PT New nucleic acid encoding nuclear mitotic appts. interacting  
PT proteins - useful for modulating cell division and proliferation and  
PT in diagnosis  
PS Claim 14; Page 28-36; 78pp; English.

CC The sequences given in W21731-32 represent fusion proteins which contain  
CC NuMA (nuclear mitotic apparatus). The fusion proteins were used in  
CC the identification of NuMA interacting proteins (NIP's) (see also  
CC W21729-30). Compounds which interfere with the interaction of NuMA  
CC with a known NIP are used to modulate cell division and/or proliferation.  
CC Ab. raised conventionally using NIP-1 or -2 as immunogen, are used to  
CC detect NIP (or their complexes) and to block their activity for  
CC diagnostic or therapeutic use, e.g. to detect defective NuMA or NIP  
CC which may be markers for aberrant (including malignant) cell growth  
CC (which can also be detected by nucleic acid sequencing). Also where  
CC malignancy is related to defects in NuMA or NIP, it can be treated by  
CC administration of the appropriate functional protein.  
SQ Sequence 2272 AA;

Query Match 5.1%; Score 161.5; DB 1; Length 2272;  
Best Local Similarity 21.0%; Pred. No. 0.00011;  
Matches 143; Conservative 93; Mismatches 285; Indels 161; Gaps 28;  
QY 33 LEALRSEHQAVLQSLSTIE-----CLQGGHHEGLVHEKARQLRRSMENIELG 81  
Db 763 LEALEKEKALEILQOQLQVANEARDSAQTSTQAKAEKLSRKVEELQACVETARQE 822  
QY 82 LSEAQVMLALASHLSTVESEKQKLRAQVRLCQENQWLDELQAGTQQL-----QRS 133  
Db 823 QHEAQAVAELE--LQLRSEQQK-ATEKERVAAQEKDQLOEQALKESLKVTKGSLEEEK 879  
QY 134 EQAVAQLEEEKHL-EFLGQLRQYDEGHTSEKEGDATKDSLDDLPNEEEDPSNGL- 191  
Db 880 RRAADALEEQRCISELKAETRSVLEQ-HKRRKE-----LEERAGRKGLE 925  
QY 192 SRQGATAAQGGYEIPAR-----LRTLH-----NLVIQYAA-QGRYE----- 228  
Db 926 ARLOQLGEAHQAETEVLRRLAEAMAAQHTAESECEQLVKEVAAWRVEDSQEEAQYG 985  
QY 229 -----VAVPLCKQALEDL-----ERTSG-----RGHPDVATM-LNILALVYRD 265  
Db 986 AMFQELMTLKEEKEKARQELQEAKEKVAGIESHSELQISRQQNELAEHLANLALQV 1045  
QY 266 QNKYKEAAHLLNDALSIRESTLGPDPHAPAAATLNNLAVLYGKRGYKEAEPLCQRALEIR 325  
Db 1046 QEKEVRAQKLADDLSTLQEK-----MAATSKVARLETILVRKAGEQETASRELVE 1097  
QY 326 EKVLTGNHPDVAKQLNNLALLCQNGKYEAVERYQYR-----ALAIYEGQ-----L 371  
Db 1098 PARAGDROPEWLEEQG-RQFCSTQAAALQAMEREAEQMGNELERLRAALMESQGOQEEER 1156  
QY 372 GPDNPNVAR--TKNNLASCYLKQGYAEAE-----TLYKEILTRAHVQE 413  
Db 1157 GQEREVARLTQERGRAQADLALEKAARAELEMLRLQNALNEQVFEFATLQEAHALTEK 1216  
QY 414 FGSVDDHDKPIWMHAEEREEMSKSR-----HHEGGTPYAEYGGWKACK 457  
Db 1217 EGKQELAKLRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHAGSGAQSEAAAG---RTE 1273  
QY 458 VSSPTVNTTLRNLGALYRR-QGKLEAAETLEEC--ALRSRRQGTDPISQT-----KVAE 508  
Db 1274 PTGPKLEALRAEVSKELEQCCQKQEQADSLERSLEAERASRAERDSALETLQGLEEKAQ 1333  
QY 509 LLGESDGRRTSQEGPGDSVKFEGGEDASVAVEWSGD---GSGTLQRSGLSKIRDVLRRS 565  
Db 1334 ELGHSQSALASAQRELAAFRTKVQDHSKAEDENKAEDEWKAQVARGRQEAERKNSL-----ISSLE 1388  
QY 566 SELLVRKLOGTEPRPSSSNMKNR 587  
Db 1389 EVSILNRQVLEKEGESKELKR 1410

RESULT 8  
R66930  
ID R66930 standard; Protein; 885 AA.  
AC R66930;  
DT 01-SEP-1995 (first entry)

DE	AMML chromosome inv(16) product.	
KW	AMML; acute myelomonocytic leukemia; chromosome-16; inversion;	
KW	inv(16); CBF-beta; CBFB gene; transcription factor; myosin; MYH11;	
KW	SMMHC.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	peptide	1..164
FT		/label= CBFB
FT	peptide	165..885
FT		/label= MYH11
PN	W09504067-A.	
PD	09-FEB-1995.	
PF	26-JUL-1994; U08530.	
PR	29-JUL-1993; US-099869.	
PA	(UNMI ) UNIV MICHIGAN.	
PA	(TEXA ) UNIV TEXAS SYSTEM.	
PI	Claxton D, Collins FS, Liu P, Siciliano MJ;	
DR	WPI; 95-082178/11.	
DR	N-PSDB; Q84589.	
PT	Novel DNA spanning the pericentric inversion of chromosome 16 -	
PT	for the screening of acute myeloid leukaemia	
PS	Claim 4; Page 34-38; 78pp; English.	
CC	PCR was performed on total cellular RNA from 5 AMML patients having	
CC	a pericentric inversion of chromosome-16, M4Eo subtype. Sequencing	
CC	showed the inv(16) fusion to comprise a sequence and the CBFB	
CC	gene, encoding a novel transcription factor, and the MYH11 gene,	
CC	encoding smooth muscle myosin heavy chain. In 1 patient, nt 1-492	
CC	of the CBFB gene were fused to nt 994 of MYH11 (shown in	
CC	Q84589; predicted aa sequence in R66930). Probes based on inv(16)	
CC	can be used for diagnosis of AMML.	
SQ	Sequence 885 AA;	
Query Match 5.1%; Score 160; DB 1; Length 885;		
Best Local Similarity 20.9%; Pred. No. 3.3e-05;		
Matches 148; Conservative 114; Mismatches 274; Indels 172; Gaps 29;		
QY	10 DEPAGHRLSQE-----EILGSTRVLSQ---LEALRSEHQAVLSQ 49	
Db	154 DRDRSHREMEAKANLDKKNQOTLEKENADLAGELRVLGQAQVEHKKKLEAQVQELQS 213	
QY	50 TIECLOQGGHEGLVHEKARQLRRSMENIELGLSEAQ---VMLA-----LASHLSTV--- 98	
Db	214 --KC-SDGERARAEELNDKVHKLQNEVESVTGMLNEAEGKAIKAKDVASLSSQLQDTQEL 270	
QY	99 --ESEKQKLR--AQVRLCQENQWLRLDELQAGTQORLORSEAVA---QLEEEKHLEFL 150	
Db	271 LQEETROKLNVSTKLRLQLEEERNLQDLDEEMAKQNLERHISTLNQLSDSKKKLQDF 330	
QY	151 GQLRQYDEGHTSEEKGDATKDSLDDLPFNEEEDPSNGLSRGQGATAAQGGYEIPAR 210	
Db	331 ASTVEALEEGKKRFQKE-----IENLTQYEEK-----AAAYDKLEKTKNR 371	
QY	211 L-RTLHNLVIQYAAQGRYEVAVPLCKQALEDLERTSGRHPDPVATMLNIALVYRDQNKY 269	
Db	372 LQQLDLDLVVDLDNQ-----RQLVSNLEKKQKQKFDQLLAEKNI--SSKYADERDR 420	
QY	270 KEA-----AHLNDALSIRESTLGPDPHPAVAATLNNLAVLYGKRGKYKEAEPL 317	
Db	421 AEAEAREKETKALSLARALEEALEAKEE--LERTNKMKAEMEDLVSSKDDVGKNVHELEK 479	
QY	318 QORALEIREKVLGTHNPVAKQLN-----NLALLCQNGKY-----EA 355	
Db	480 SKRALETQMEEMKTQLEEELEDELOQASEDAKLRLEVNMQAL---KGQFERDLQARDEQNEE 536	
QY	356 VERYQRALAIYEGQLGPDNPVARTKNNLASCYLK-----QGYAE 398	
Db	537 KRRLOLQRLHEYTELEDE-----RNERALAAAKKKLEGLDKLELQADSIAIKGREEAI 591	
QY	399 ETLYKEILTRAHVQFGSVDDHK-----PIWMHAEEREEMSKSRHHEGGTPYAEYGGWYK 454	
Db	592 KQLRK---LQQMCKDFQRELEDARASRDEIFATAKENEKKAKSLEAD----- 635	
QY	455 ACKVSSPTVNTTLNLGAL--YRRQKLEAAETLEECAL----RSRROGTDPISQTKVAE 508	
Db	636 -----LMQLQEDLAAARARKQADLEKEBELAEELASSLSGRNALQDEKRRLEARIAQ 687	
QY	509 LLGESDGRRTSQEGPGDSVKFEGGEDASVAVWMSGSGTTLQSRGSLGKIRDVLRRSSEL 568	
Db	688 LEEELQGNMEAMSDRVYRKATQQAELSNELATERS-TAQKNES---ARQQLERQNK 743	
QY	569 LVRKLGTEPRPSSSNMKRAASLN-YLNQPSAAPLQVSRGLSASTMDL 615	
Db	744 LRSKLHEMEGAVKSKFKSTIAALEAKIAQLEEQVEQEAQKQATKSL 791	
RESULT 9		
Y07058		
ID	Y07058 standard; Protein; 543 AA.	
AC	Y07058;	
DT	02-JUL-1999 (first entry)	
DE	Renal cancer associated antigen precursor sequence.	
KW	Cancer associated antigen; diagnosis; research; treatment; human;	
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
KW	prostate cancer.	
OS	Homo sapiens.	
PN	W09904265-A2.	
PD	28-JAN-1999.	
PF	15-JUL-1998; U14679.	
PR	22-JUN-1998; US-102322.	
PR	17-JUL-1997; US-896164.	
PR	10-OCT-1997; US-061599.	
PR	10-OCT-1997; US-061765.	
PR	10-OCT-1997; US-948705.	
PR	11-OCT-1997; GB-021697.	
PA	(LUDW-) LUDWIG INST CANCER RES.	
PI	Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,	
PI	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,	
PI	Tureci O;	
DR	WPI; 99-132448/11.	
PT	New isolated cancer associated nucleic acids and polypeptides -	
PT	isolated using sera from cancer patients, used to develop products	
PT	for the diagnosis, monitoring or treatment of cancers	
PS	Disclosure; Page 448-449; 787pp; English.	
CC	The invention relates to a method for diagnosing a disorder characterised	
CC	by expression of a human cancer associated antigen precursor coded for by	
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a	
CC	biological sample isolated from a subject with an agent that specifically	
CC	binds to the NAM, an expression product or a fragment of an expression	
CC	product complexed with an HLA molecule; and (b) determining the	
CC	interaction between the agent and the NAM or the expression product as a	
CC	determination of the disorder. The products and methods can be used in	
CC	the diagnosis, monitoring, research, or treatment of conditions	
CC	characterised by the expression of various cancer associated antigens.	
CC	The invention provides nucleic acid sequences and encoded polypeptides	
CC	which are cancer associated antigen precursors expressed in human breast	
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and	
CC	lung cancer.	
SQ	Sequence 543 AA;	
Query Match 5.0%; Score 159.5; DB 1; Length 543;		
Best Local Similarity 21.4%; Pred. No. 1.7e-05;		
Matches 86; Conservative 61; Mismatches 137; Indels 117; Gaps 17;		
QY	38 SEHQAVLSQTSQIECLOQGGHEGLVHEKAR-QLRRRSMENIELGLSEAQVMLALASHLS 96	
Db	76 SRKAAALEFLNRFEEAKRT--YEGLKHEANNPQLKEGLQNMEARLAERKFM----- 125	
QY	97 TVESEKQLRAQVRRLCQENQWLRLDELQAGTQORLORSEQAQVALEEEKHLEFLGQLRQY 156	
Db	126 -----NPF---NMPNLYQKLE-SDPRTRLSDPTYRELIEQLRNK 162	
QY	157 DEDGHTS-----BEKGDAT-----KDSLDDLPNNEE--E 185	
Db	163 PSDLGTGLQDPRIINTLSVLLGVLDGSMDEEEIATPPPPPPPKKTKPE--PMEEDLPE 220	

```
QY 186 DPSNGLSRGQ-GATAAQQGGYEI-----PARLRTLHNLVIOYAAQGRYEVAVP 232
Db 221 NKKQALKEKELGNDAYKKKDFDTALKHYDKAKELDPTNMTYITNOAAVYFEKGDYNKCRE 280
QY 233 LCKQALDELRTSGRHPDVATMLNILALVYRDONKYKEAAHLLNDALSIRESTLGPDP 292
Db 281 LCEKAIE-VGRENREDYQIAKAYARIGNSYFKEEKYKDAIHFNYSLSAEHRT--PD-- 334
QY 293 AVAATLNNLAVLYGKRGYKEAEPLCQRALEI--REKVLGTNHPDVAKQLNNLALLCQ 350
Db 335 -----VL-----KCKQQAEEKILKEQERLAYINPDLALEEKNKGNECFQK 373
QY 351 GKYEAVERYQORALAIYEGQLGPDNPNVARTKNNLASCYLK 391
Db 374 GDYPQAMKHYTEAI-----KRNPKDAKLYSNRAACYTK 406
```

```
RESULT 10
W19540
ID W19540 standard; Protein; 1325 AA.
AC W19540;
DT 16-SEP-1997 (first entry)
DE Male-enhanced antigen-2.
KW Mouse; MEA-2; detecting mutation.
OS Mus musculus domesticus.
FH Key Location/Qualifiers
FT misc_difference 305..320
FT /note= "Not shown in the specification"
```

```
PN J09121869-A.
PD 13-MAY-1997.
PF 07-NOV-1995; 311638.
PR 07-NOV-1995; JP-311638.
PA (ITOH-) ITO HAM KK.
DR WPI; 97-314229/29.
DR N-PSDB; T74034.
PT Male-enhanced antigen Mea-2 gene - especially from mouse, useful for
PT detecting mutation(s)
PS Claim 8; Page 9-10; 13pp; Japanese.
CC The present sequence represents male-enhanced antigen-2 (MEA-2), which
CC has been derived from a domestic mouse. The polynucleotide encoding
CC the protein can be used for the detection of mutations affecting the
CC MEA-2 gene.
SQ Sequence 1325 AA;
```

Query Match 4.8%; Score 152.5; DB 1; Length 1325;  
Best Local Similarity 21.3%; Pred. No. 0.00028;  
Matches 134; Conservative 104; Mismatches 219; Indels 173; Gaps 30;

```
QY 7 GQRDE---PAGHRLSQEEILGSTRLV---SQGLEALRSEHQAVLQSLSQTIETCLOQGGH 59
Db 513 GEREQLQKVADAAASLEQQLEQVKLTFLFORDQQLAALQQEHLDVIKLTSTQEALQAKGQ 572
QY 60 EEGLVHEKARQLRRSMENIELG-----LSEAQVLMALASHLSTVSESEKQLRAQV 109
Db 573 SLDDLHTRYDELOARLEELQREADSREDAIHFLQNEKIVLEVA--LQSAKSDKEELDRGA 630
QY 110 RRLCOEN-----QWLRLDELACTQORLQORSEQAVALQEEKKHL--EFLGQ---LRQYD 157
Db 631 RRLEEDTEETSGLLEQLRQDLAVKSNQVEHLQQETATLRKQMKVKEQFVQQKVMVEAYR 690
QY 158 EDGHTSEE--KEGDATEKDSLDLFPNEEEDPSNGLSRGQATAAQQGGYEIP-ARLRTL 214
Db 691 RDATSKDQLINELKATKKRLD-----SEMKELRQELIKIQG----EKKTVEVHSRLQKD 741
QY 215 HNLVIOYAAQ-----GRYEVAVPL-----CKQALEDLERTS 245
Db 742 MSLVHQMAELEGHLQSVQKERDEMEIHLQSLKFDKEQMIALTEANETLKKQIBELQOE 801
QY 246 GRGHPDVATMLNILALVYRDQNK-----YKEAAHLN-----DALSIRE----- 284
Db 802 KKAITEQKQMKRGLSDLTSAQKEMKTKHAYENAVSILSRRLQELALASKEATDAELNQL 861
```

```
QY 285 ---STLGPDPHPAATAATLNNLAVLYGKRGYK---EAE-----PLCQRALEIREKVLGTN 332
Db 862 RAQSTGGSDPVLHEKIRALELEVELQNVGQSILLEKELEQEVITMTSQEELSESREKVLELE 921
QY 333 HP-----DVAKQLNNLALLCQNOGKYEAVERYQORALAIYEGQL-GPDNPNVARTKNN-- 384
Db 922 DELQESGRFRKIKRL----EESNKKLAELEHER-----GKLTGLGQSNAAALREHNSI 971
QY 385 LASCYLKQGYAEAEITLYKEILTRAHVQEFSGVDDDKPIWMHAE---EREEMSKSRHHE 441
Db 972 LETALAKR---EADLVQLNLQVAVLQORKEEEDROMKQLVQALQVLSLEKEME----- 1021
QY 442 GGTPYAEYGGWYKACKVSSPTVNTTLRNLGALYRQGGKLEA-----AETLEECALRS 493
Db 1022 -----VNSLKEQMAA-----ARIEAGHNRHFRKAAATLELSEVKK 1055
```

```
QY 494 RROGTDPISQTKVAEL--LGESDGRRTSQE 521
Db 1056 ELQAKEHLVQTLQAEVDELQIQDGKH-SQE 1084
```

```
RESULT 11
W94391
ID W94391 standard; Protein; 1325 AA.
AC W94391;
DT 14-APR-1999 (first entry)
DE Mouse male enhanced antigen 2.
KW Mouse; male enhanced antigen 2; Mea-2; Mus musculus domesticus;
KW spermatogenesis; regulation; contraceptive; sterile; inhibition.
OS Mus sp.
PN J11018622-A.
PD 26-JAN-1999.
PF 04-JUL-1997; 179490.
PR 04-JUL-1997; JP-179490.
PA (ITOH-) ITO HAM KK.
DR WPI; 99-160962/14.
DR N-PSDB; X04132.
PT Regulation of spermatogenesis using Mea-2 gene information - using
PT anti-sense oligo- or poly:nucleotide(s), used for production of
PT contraceptives
PS Claim 4; Page 8-12; 27pp; Japanese.
CC The present sequence represents mouse male enhanced antigen 2 (Mea-2).
CC The present invention describes the regulation of spermatogenesis by
CC using Mea-2 information. A non-human living organism can have its
CC spermatogenesis inhibited by breakage of the whole or part of the Mea-2
CC gene. Also described are: (1) the creation of the whole or part of the
CC inhibited organism; (2) a drug composition containing an oligonucleotide
CC or polynucleotide containing base sequences that pair with at least part
CC of the Mea-2 gene and are able to inhibit the expression of Mea-2 gene;
CC and (3) the creation of an aimed gene-possessing organism using the
CC spermatogenesis inhibited organism. The organism is useful for producing
CC contraceptive drugs.
SQ Sequence 1325 AA;
```

Query Match 4.8%; Score 152.5; DB 1; Length 1325;  
Best Local Similarity 21.3%; Pred. No. 0.00028;  
Matches 134; Conservative 104; Mismatches 219; Indels 173; Gaps 30;

```
QY 7 GQRDE---PAGHRLSQEEILGSTRLV---SQGLEALRSEHQAVLQSLSQTIETCLOQGGH 59
Db 513 GEREQLQKVADAAASLEQQLEQVKLTFLFORDQQLAALQQEHLDVIKLTSTQEALQAKGQ 572
QY 60 EEGLVHEKARQLRRSMENIELG-----LSEAQVLMALASHLSTVSESEKQLRAQV 109
Db 573 SLDDLHTRYDELOARLEELQREADSREDAIHFLQNEKIVLEVA--LQSAKSDKEELDRGA 630
QY 110 RRLCOEN-----QWLRLDELACTQORLQORSEQAVALQEEKKHL--EFLGQ---LRQYD 157
Db 631 RRLEEDTEETSGLLEQLRQDLAVKSNQVEHLQQETATLRKQMKVKEQFVQQKVMVEAYR 690
QY 158 EDGHTSEE--KEGDATEKDSLDLFPNEEEDPSNGLSRGQATAAQQGGYEIP-ARLRTL 214
```



Db 691 RDATSKDQLINELKATKKRLD-----SEMKELRQELIKLQG-----EKKTVEVEHSRLQKD 741  
QY 215 HNLVIQYAAQ-----GRYEVAVPL-----CKQALEDLERTS 245  
Db 742 MSLVHQMAELEGHLQSVQKERDEMEIHLQSLKFDKEQMIALTEANETLKKQIEELQOEA 801  
QY 246 GRGHPDVATMLNILALVYRDQNK-----YKEAAHLN-----DALSIRE----- 284  
Db 802 KKAITEQKQMKRGLSDLTSAQKEMKTKHKAYENAVSILSRRLQEALASKEATDAELNQL 861  
QY 285 ---STLGPDPHPAATLNNLAVLYGKRGKYK---EAE-----PLCQRALEIREKVLGTN 332  
Db 862 RAQSTGGSSDPVLHEKIRALELEVELQNVGQSKILLEKELOEVITWTSQELEESREKVLLE 921  
QY 333 HP-----DVAKQLNLLCQNGQKYEAVERYQORALAIYEGQL-GPDNPNVARTKNN-- 384  
Db 922 DELQESRGFRKIKRL-----EESNKKLALELEHER-----GKLTGLGQSNAALREHNSI 971  
QY 385 LASCYLKQGYAEATLYKEILTRAHVQEFSGVDDHDKPIWMHAE---EREEMSKSRHHE 441  
Db 972 LETALAKR-----EADLVQLNLQVQAVLQVRKEEDRQMQLVQALQVLSKEKME----- 1021  
QY 442 GGTPYAEYGGWKACKVSSPTVNTTLRNLGALYRRQKLEA-----AETLEECALRS 493  
Db 1022 -----VNSLKEQMAA-----ARIEAGHNRHRHFKAAATLELSEVKK 1055  
QY 494 RRQGTDPISQTKVAEL--LGESDGRRTSQE 521  
Db 1056 ELQAKEHLVQTLQAEVDELQIQDGKH-SQE 1084

RESULT 12

W82500  
ID W82500 standard; Protein; 920 AA.  
AC W82500;  
DT 01-FEB-1999 (first entry)  
DE Human OGT protein.  
KW OGT; O-linked GlcNAc transferase; uridine; transferase; human; tumour;  
KW diphospho-N-acetylglucosamine; polypeptide beta-N-acetylglucosaminyl;  
KW predisposition; type II diabetes; hyperglycaemia; Alzheimer's disease;  
KW metastasis; diagnosis; inhibitor; treatment; diabetes mellitus.  
OS Homo sapiens.  
PN WO9844123-A2.  
PD 08-OCT-1998.  
PF 27-MAR-1998; U06101.  
PR 31-MAR-1997; US-042270.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI Hanover JA, Lubas W;  
DR WPI; 98-557118/47.  
DR N-PSDB; V69301.  
PT Protein exhibiting O-linked GlcNAc transferase activity, OGT -  
PT useful, e.g. to assess predisposition to type II diabetes or  
PT Alzheimer's or metastatic potential of tumours, and to identify  
PT inhibitors  
PS Claim 7; Page 31-33; 56pp; English.  
CC This sequence represents a novel human O-linked GlcNAc transferase, OGT  
CC protein (also known as uridine diphospho-N-acetylglucosamine:  
CC polypeptide beta -N-acetylglucosaminyl transferase). This protein is  
CC useful to assess predisposition toward type II diabetes in patients  
CC suspected of having hyperglycaemia that could evolve into this disease,  
CC by assaying OGT activity in red blood cells. It can also be used to  
CC assess predisposition toward Alzheimer's disease, to assess the  
CC metastatic potential of tumours and to diagnose a tumour with metastatic  
CC potential. OGT can also be used to identify OGT inhibitors, especially in  
CC high-throughput assays, useful, e.g. in the treatment of diabetes  
CC mellitus, tumour-derived diseases and Alzheimer's disease.  
SQ Sequence 920 AA;

Query Match 4.8%; Score 151.5; DB 1; Length 920;  
Best Local Similarity 24.2%; Pred. No. 0.00019;  
Matches 80; Conservative 42; Mismatches 112; Indels 97; Gaps 14;

QY 207 IPARLRTLH-----NLVIQAAQGRYEVAVPLCKQALEDLERTSGRHPDVATML 256  
Db 119 VAAYLRALSLSPNHAVVHGNLACVYYEQGLDLAIDTYRRAIE-----LQPHFPDAY 170  
QY 257 NILALVYRDQNKYKERAAHLNDALSIRESTLGPDPHPAATLNNLAVLYGKRGKYKEAEP 316  
Db 171 CNLANALKEKGSVAEEDCYNTAL-----RLCPH---ADSLNNLANIKREQGNIEAVR 222  
QY 317 LCQRALEIREKVLGTNHPDVAKQLNLLCQNGQKYEAVERYQORALAI----- 366  
Db 223 LYRKALEV-----FPEFAAAHSNLASVLQQQGLQEALMHYKEAIRISPTFADAYSN 274  
QY 367 -----YEQQLG-----PDNPNVARTKNNLASCYLKQGYAEATLYKEIL---- 406  
Db 275 MGNTLKEMQDVQGALQCYTRAIQINPAFADAHSNLASIHKDSGNIPETAISYRTALKLP 334  
QY 407 -----TRAHVQEFSGVDDHDKPIWMHAEER-----EEMSKSR-----HHEGGT 444  
Db 335 DFPDAYCNLAHCLQI--VCD-----WTDYDERMKKLVLSIVADQLEKNRPLPSVPHHSMLY 387  
QY 445 PYAEYGGWKACKVSSPTVNTTLRNLGALYR 475  
Db 388 PLSH--GFRKA--IAERHGNLCLDKINVLHK 414

RESULT 13

R57365  
ID R57365 standard; Protein; 955 AA.  
AC R57365;  
DT 08-MAR-1995 (first entry)  
DE K39 polypeptide of leishmania chagasi.  
KW Parasite; leishmaniasis; infection; repeat unit; vaccine; antigen;  
KW Leishmania chagasi.  
OS Leishmania chagasi.  
PN WO9416331-A.  
PD 21-JUL-1994.  
PF 10-JAN-1994; U00324.  
PR 15-JAN-1993; US-006676.  
PA (IASY-) IASYS CORP.  
PI Reed SG;  
DR WPI; 94-249402/30.  
DR N-PSDB; Q70152.  
PT Diagnosis of Leishmaniasis - by determining the presence of  
PT antibodies that bind to a K39 repeat unit antigen  
PS Disclosure; Page 12-15; 28pp; English.  
CC The K39 polypeptide comprises a number of repeated units (described  
CC in R57366). Detection of antibodies directed against this repeated  
CC unit in a patients sample is indicative of leishmaniasis. The  
CC antigenic repeat unit can itself be used as a vaccine to protect  
CC against infection by a leishmania parasite.  
SQ Sequence 955 AA;

Query Match 4.7%; Score 149.5; DB 1; Length 955;  
Best Local Similarity 20.9%; Pred. No. 0.0003;  
Matches 129; Conservative 92; Mismatches 274; Indels 121; Gaps 25;

QY 49 QTIECLQQGGHEGLVH-----EKARQLRRSMENIELGLS--EAQVMLALASHLS 96  
Db 378 ETLSTLRYASRARDIVNVAQVNEPPRRRIRFELEEQMEDMRQAMAGGDPAYVSELKKKLA 437  
QY 97 TVESEKQKLRAQVRRLCQE---NQWLRLDELQAGTQORLQRSEQAVQAQLEEKHLEFLGQL 153  
Db 438 LLESEAQKRAADLQALEREREHQNQVQERLLRATEAEKSELESRAALQE-----EMTATR 492  
QY 154 RQYDE-DGHTSEEKGDATKDSLDDLPFNEEEDPSNGLSRGQGATAAQGGYEIPARLR 212  
Db 493 ROADKMQALNLRUKEEQARKER--ELKEMAKKDA--LSK-----VRRRKDAEIASERE 543  
QY 213 TLHNLVIQYAAQGR-YEVAVPL-----CKQALEDLERTSGRHPDVATMLNILALVYR 264  
Db 544 KLESTVAQLEREQEREREVALDALQTHQRKLQEALESSERTAAE-----R 587



QY	6	LGQRDEPAGHRLSQE--EILGSTRLVSQGLEALRS-----EHQAVLQSLSQTIETCICQQG	57
Db	951	LEKESQDSKSLSEDESQETFGP--LEKENAESLRSLAGODQOEQKLEQETOQTTLRAV---	1005
QY	58	GHEEGLVH--EKA-----RQLRRSM--ENIELGLSEAOVMLALASHLST-----	97
Db	1006	GNEQMAVSPPEKVDPPELKPPLGNDQEIARSLGKXENQESLSVLKKGKGTYSKSLTEITIEP	1065
QY	98	VESEKQKLRAQVRRRLCOENQWL-----RDELAGTQQLRQORSEQAQAVALKEBKHH	146
Db	1066	LETAEEDELRKRSIDTQEPWLSTEVARETVPEPPDEPPGSLGSDVENRETTLSLEKESQE	1125
QY	147	LEFLG-----QLROYDEDGHTSEEKEG-DATKDSLDDLPNEEEEDPSNGLSR	193
Db	1126	LSSLGKNWVETRVEDSQCLQVEEGLQEEHQHQSLEREVKQELPSSGNQQRWEDVVECKAV	1185
QY	194	QGATATA-----QQGGYEIPARLTLNLNVIQYAAQGR--YEVAVPLC	234
Db	1186	GQEAPLATTVGVTEDKAELHLRGQGGEEAAAEGELLQDIVGEAWSLGSSEPKQRYP--	1243
QY	235	KOALEDLERTSGRHPDVATMLNIALVYRDQNKYK-----EAAHLNDALS	281
Db	1244	AEALDNLE---GGALEVPVQAQSMPEVTERDEDRQAQAGEQDSIEVTGLGEAA--RTGLE	1296
QY	282	IRESTLGPDPH-----AVAATLNNLAV-----LYGKRKGYKEAEPLCQRALEIREK	327
Db	1297	LEQEVVGLDEPRHFAREEAIPDSLGEESVKAKIAQGLEPGPKPEKAGALDSGILEL---	1353
QY	328	VLGTNHPDVAKQNLNALLCQNGQKYEAVERYIQRALAI---YEQQLGPDNPVARTKN	383
Db	1354	-----PKTSS-----ALECQGHSESESMEGWEEEEEASLETSDHEGSDAP-QPRPPTEE	1402
QY	384	NLASCYLKQKGYAEATLYK-----EILTRAH-----VQEFSGVDDDHKPIWMHAE	430
Db	1403	DEGA---QAALTA PGPKLLEPCSPILPILTDHAHELQPOAEGIQEAGWQPEAGSEALERVEN	1459
QY	431	REEMSKSRHHEG-----GTPYAEYGGWYKACKVS-----	459
Db	1460	EPEFGLGEIPEGLQDWEEGRESEADDLGETLPDSTPL---GLYLRSPASPKWDLAGEQ	1515
QY	460	--SPTVNTLRLNLG-ALYRRQCKLEAAETLEECALRSRRQGTD-----	499
Db	1516	RLSPQGDAGKEDWGPVAPAAQGLSGPPEEEEE-----QGHGSDLSSEEFEDLGTEASLLP	1570
QY	500	-----PISQTKVAELLGESDG-----RRTSQEGPGDSVKFEGGEDASVAVEM	541
Db	1571	GVPKEVADHVQVPPVLQPCADWQGGESDGFADBEESGEEGEEDADEEGAESGA---QW	1627
QY	542	SGDGSGLTLRSGSLGKIRDVLRRSELL---VRKLQGTETPRPSSNMKRAASLNYLNQP	597
Db	1628	WGSQA-----SGGCKVKQDIAQRGDPVQESVGVSGLWDGLRGAANV---PALEMVSQD	1679
QY	598	SAAPLQVSRGLSAS	611
Db	1680	SAEPSGSESESAS	1693



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2000, 04:28:15 ; Search time 58.44 Seconds  
(without alignments)  
162.367 Million cell updates/sec

Title: US-09-036-614A-1  
Perfect score: 3161  
Sequence: 1 MSGLVLQRPAGRLSQE.....APLQVSRGLSASTMDLSSSS 619

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues  
Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	167.5	5.3	2101	1 US-08-466-390-4	Sequence 4, Appli
2	167.5	5.3	2101	1 US-08-470-950-4	Sequence 4, Appli
3	167.5	5.3	2101	1 US-08-467-781-4	Sequence 4, Appli
4	167.5	5.3	2101	1 US-08-195-487-4	Sequence 4, Appli
5	167.5	5.3	2101	2 US-08-483-924-4	Sequence 4, Appli
6	167.5	5.3	2101	4 PCT-US93-06160-4	Sequence 4, Appli
7	163	5.2	981	2 US-08-649-046-2	Sequence 2, Appli
8	160	5.1	885	2 US-08-533-306A-4	Sequence 4, Appli
9	160	5.1	885	2 US-08-742-923A-4	Sequence 4, Appli
10	149.5	4.7	955	1 US-08-006-676B-1	Sequence 1, Appli
11	149.5	4.7	955	1 US-08-282-845-2	Sequence 1, Appli
12	149.5	4.7	955	2 US-08-428-414A-3	Sequence 3, Appli
13	149.5	4.7	955	4 PCT-US94-00324-1	Sequence 1, Appli
14	149.5	4.7	1805	1 US-07-853-913-2	Sequence 1, Appli
15	148	4.7	2482	1 US-08-328-254-6	Sequence 2, Appli
16	148	4.7	3248	1 US-08-353-700-1	Sequence 1, Appli
17	148	4.7	3248	4 PCT-US95-16216-1	Sequence 1, Appli
18	147	4.7	2154	2 US-08-841-349-4	Sequence 4, Appli
19	144	4.6	816	2 US-08-533-306A-6	Sequence 6, Appli
20	144	4.6	816	2 US-08-742-923A-6	Sequence 6, Appli
21	139	4.4	477	1 US-08-402-217A-3	Sequence 3, Appli
22	139	4.4	477	1 US-08-700-178-3	Sequence 3, Appli
23	139	4.4	477	3 US-08-995-654-3	Sequence 3, Appli
24	137	4.3	546	2 US-08-533-669A-2	Sequence 2, Appli
25	136	4.3	576	2 US-08-533-306A-2	Sequence 2, Appli
26	136	4.3	576	2 US-08-742-923A-2	Sequence 2, Appli
27	135.5	4.3	1354	3 US-08-685-871-2	Sequence 2, Appli
28	129	4.1	1618	1 US-07-853-913-4	Sequence 4, Appli

29	128.5	4.1	1388	2 US-08-685-576-1	Sequence 1, Appli
30	127	4.0	3031	1 US-07-689-008-2	Sequence 2, Appli
31	125.5	4.0	1093	4 PCT-US93-03077-1	Sequence 1, Appli
32	125.5	4.0	1898	1 US-08-056-200-94	Sequence 94, Appli
33	125.5	4.0	1898	2 US-08-800-644-94	Sequence 94, Appli
34	124.5	3.9	683	5 5210183-3	Patent No. 5210183
35	123	3.9	1312	2 US-08-592-126-148	Sequence 148, App
36	122.5	3.9	1388	2 US-08-685-576-4	Sequence 4, Appli
37	122	3.9	1566	2 US-08-687-956A-23	Sequence 23, Appli
38	121	3.8	1312	2 US-08-687-080-51	Sequence 51, Appli
39	116.5	3.7	645	2 US-08-592-126-144	Sequence 144, App
40	116.5	3.7	645	2 US-08-687-080-47	Sequence 47, Appli
41	116	3.7	2842	1 US-07-741-940-7	Sequence 7, Appli
42	116	3.7	2842	1 US-08-289-548A-7	Sequence 7, Appli
43	116	3.7	2842	1 US-08-452-654-7	Sequence 7, Appli
44	116	3.7	2843	1 US-08-452-655B-2	Sequence 2, Appli
45	116	3.7	2843	1 US-08-452-655B-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-08-466-390-4  
; Sequence 4, Application US/08466390  
; Patent No. 5686562  
; GENERAL INFORMATION:  
; APPLICANT: TOUKATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,390  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-466-390-4

Query Match	5.3%;	Score 167.5;	DB 1;	Length 2101;
Best Local Similarity	20.5%;	Pred. No. 5.9e-06;		
Matches	156;	Conservative 106;	Mismatches 272;	Gaps 32;
QY	15	HRLSQEEILGSTRVLSQGLEA-----LRSEHQAVLQSLSTI-----ECLQ---	55	
Db	750	HKRERKE-LDEERAGRKGLEARLLQLGEAHQAETEVLRLRELAFAEAMAAQHTAESECEQLVK	808	
QY	56	-----QGGHEE-----GLVHEKARQLRRSMENIELGLSEAQVMLA-LASH-----LS	96	

```
Db 809 EVAAWRDGYEDSQEEAAQYGAMFQEQMLTKKECEKARQELQEAKEKVAGIESHSELQIS 868
QY 97 TVESEKQKLRQVRRLCQE-----NQWLDELQAGTQQLRQSEQAQAQLE----- 141
Db 869 RQONKLAELHANLARALQQVQVEKVRQAQLADDLSTLQEKMAATSKVEVARLETTLVRKAGE 928
QY 142 -EEKKHLEFLGQ-LRQYDEDGHTSEEKGD---ATKDSLDDLPNEEEEDPSNGLSRGQG 196
Db 929 QQETASRELKVEPARAGDRQPEWLEEQQGRQFCSTQAALQAM--ERAEQMGNELERLRA 986
QY 197 ATAAQQGGYEIPARLRTLHNLVQYAAQGRYEVAVPLCKQALEDLERTSGRHPDVATML 256
Db 987 ALMESQGGQ-----QQEERGQOE-----REVARLTQERGRAQADLALEK 1024
QY 257 NILA-LVYRDQNKYKE-----AAHLNDALSIRESTLGPDPHPAVAATLNNLAVLYG-K 307
Db 1025 AARAELEMLRLQNALNEQORVEFATLQELAHALTEKE---GKD-----QELAKLRGLE 1073
QY 308 RGYKKEAEPLCQRALEIREKV-----LGTNHPDVAK-----QLNNLA 344
Db 1074 AAQIKELEELRQTVKQKEQLAKKEKEHSGSGAQSEAAGRTEPTGPKLEALRAEVSKLE 1133
QY 345 LLCQNGKY-EAVERYIQ-----RALAIYEGQ-----LGPDPNPVARTKNNLA 386
Db 1134 QQCQKQEQADSLERSLEAERASRAERDSALETLQGLEEKAQELGHSQSALASARELA 1193
QY 387 SCYLKQKGKYAEATLYK-----RQALAIYEGQ-----EILTRAHVQEFSGVDDDHK 422
Db 1194 AFRTKVQDHSKADEWKAQVARGRQEAERKNSLISSEEEVSILNRQVLEKEGESKELKR 1253
QY 423 PIWMHAEEREEMSKSRHHEGGTPYAEYGGWKACKVSSPTVNTTLRNLGALYRRQKLEA 482
Db 1254 LVMAESEKSKLEES-----CACCRRQRPATVPPELQNAALLCGRRRCRAG 1298
QY 483 AETLEECALRSRRQGTDPISQTKVAELLGESDGRRTSQEGPGDSVKFEGGEDASVAVENS 542
Db 1299 REA-EKQVASENLRLQELTQSAERAEELG-----QELKAWQEKFFQKEQALSTLQLE 1349
QY 543 GDGSGTL-----QKSGSLGKIR-DVLRSSSEL--L 569
Db 1350 HTSTQALVSELLPAKHLCCQLQAEQAAAEKRRHREELEQSKQAAGGLRAELLRQRELDEL 1409
QY 570 VRKLQGTETPRPSSNNMKRAASLNLYNQPSAAPLQVSRGLSA 610
Db 1410 IPLRQKVAEQERTAAQLRAEKASYAEQLSM--LKKAHGLLA 1448
```

RESULT 2

```
US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,950
; FILING DATE: 06-JUN-1995
```

```
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-950-4
```

```
Query Match 5.3%; Score 167.5; DB 1; Length 2101;
Best Local Similarity 20.5%; Pred. No. 5.9e-06;
Matches 156; Conservative 106; Mismatches 272; Indels 227; Gaps 32;

QY 15 HRLSQEEILGSTRLVSQGLEA---LRSEHQAVLQSLSQTI-----ECLQ--- 55
Db 750 HKRERKE-LEERAGRKGLEARLLQLGGEAQAETEVLRRLEAFAMAAQHTAESECEQLVK 808
QY 56 ----QGGHEE-----GLVHEKAPQLRRSMENIELGLSEAQVMLA-LASH----LS 96
Db 809 EVAAWRDGYEDSQEEAAQYGAMFQEQMLTKKECEKARQELQEAKEKVAGIESHSELQIS 868
QY 97 TVESEKQKLRQVRRLCQE-----NQWLDELQAGTQQLRQSEQAQAQLE----- 141
Db 869 RQONKLAELHANLARALQQVQVEKVRQAQLADDLSTLQEKMAATSKVEVARLETTLVRKAGE 928
QY 142 -EEKKHLEFLGQ-LRQYDEDGHTSEEKGD---ATKDSLDDLPNEEEEDPSNGLSRGQG 196
Db 929 QQETASRELKVEPARAGDRQPEWLEEQQGRQFCSTQAALQAM--ERAEQMGNELERLRA 986
QY 197 ATAAQQGGYEIPARLRTLHNLVQYAAQGRYEVAVPLCKQALEDLERTSGRHPDVATML 256
Db 987 ALMESQGGQ-----QQEERGQOE-----REVARLTQERGRAQADLALEK 1024
QY 257 NILA-LVYRDQNKYKE-----AAHLNDALSIRESTLGPDPHPAVAATLNNLAVLYG-K 307
Db 1025 AARAELEMLRLQNALNEQORVEFATLQELAHALTEKE---GKD-----QELAKLRGLE 1073
QY 308 RGYKKEAEPLCQRALEIREKV-----LGTNHPDVAK-----QLNNLA 344
Db 1074 AAQIKELEELRQTVKQKEQLAKKEKEHSGSGAQSEAAGRTEPTGPKLEALRAEVSKLE 1133
QY 345 LLCQNGKY-EAVERYIQ-----RALAIYEGQ-----LGPDPNPVARTKNNLA 386
Db 1134 QQCQKQEQADSLERSLEAERASRAERDSALETLQGLEEKAQELGHSQSALASARELA 1193
QY 387 SCYLKQKGKYAEATLYK-----EILTRAHVQEFSGVDDDHK 422
Db 1194 AFRTKVQDHSKADEWKAQVARGRQEAERKNSLISSEEEVSILNRQVLEKEGESKELKR 1253
QY 423 PIWMHAEEREEMSKSRHHEGGTPYAEYGGWKACKVSSPTVNTTLRNLGALYRRQKLEA 482
Db 1254 LVMAESEKSKLEES-----CACCRRQRPATVPPELQNAALLCGRRRCRAG 1298
QY 483 AETLEECALRSRRQGTDPISQTKVAELLGESDGRRTSQEGPGDSVKFEGGEDASVAVENS 542
Db 1299 REA-EKQVASENLRLQELTQSAERAEELG-----QELKAWQEKFFQKEQALSTLQLE 1349
QY 543 GDGSGTL-----QKSGSLGKIR-DVLRSSSEL--L 569
Db 1350 HTSTQALVSELLPAKHLCCQLQAEQAAAEKRRHREELEQSKQAAGGLRAELLRQRELDEL 1409
QY 570 VRKLQGTETPRPSSNNMKRAASLNLYNQPSAAPLQVSRGLSA 610
Db 1410 IPLRQKVAEQERTAAQLRAEKASYAEQLSM--LKKAHGLLA 1448
```





```
Best Local Similarity 20.5%; Pred. No. 5.9e-06;
Matches 156; Conservative 106; Mismatches 272; Indels 227; Gaps 32;

QY 15 HRLSQEELGSTRLVSQGLEA-----LRSEHQAVLQSLSTI-----ECLQ--- 55
Db 750 HKRERKE-LEERAGRKGLARLLQLGEAHQAETEVLRRLEAEAMAAQHTAESECEQLVK 808

QY 56 -----QGGHEE-----GLVHEKARQLRRSMENIELGLSEAQVMLA-LASH-----LS 96
Db 809 EVAAWRDGYEDSQEEAQYGAMFQEQMLTLKEEKEKARQELQEAKEKVAGIESHSELQIS 868

QY 97 TVESEKQKLRAQVRRLCQE-----NQWLRDELQAGTQORLQSEQAQAOLE----- 141
Db 869 RQONKLAELHANLARALQQVQVEKVRQAQLADDLSTLQEKMAATSKAVARLETLVRKAGE 928

QY 142 -EEKKHLEFLGQ-LROYDEDGHTSEKEGD---ATKDSLDDLPFNEEEEDPSNGLSRGQG 196
Db 929 QQETASRELKVPARAGDRQPEWLEEQGGRQFCSTQAALQAM--ERAEQMGNELERLRA 986

QY 197 ATAAQQGGYEIPARLRTLHNLVQYAAQGRYEVAVPLCKQALDLELRTSGRHPDVATML 256
Db 987 ALMESQGG-----QOERGGQE-----REVARTQERGRAQADLAEK 1024

QY 257 NILA-LVYRDQNKYKE-----AAHLNDALSIRESTLGPDPHPAVAATLNNLAVLYG-K 307
Db 1025 AARAELEMLQNALNEQORVEFATLQELAHALTEKE---GKD-----QELAKLRGLE 1073

QY 308 RGKYKEAEPLCQRALEIREKV-----LGTNHPDVAK-----QLNNLA 344
Db 1074 AAQIKELEELRQTVKQLKEQLAKKEKEHSGSGAQSEAAGRTPTGPKLEALRAEVSKLE 1133

QY 345 LLCQNGKY-EAVERYIQ-----RALAIYEGQ-----LGPDPNPVARTKNNLA 386
Db 1134 QQCQKQEQADSLERSLEAERASRAERDSALETLQGLEKAQELGHSQSALASARELA 1193

QY 387 SCYLKQGYAEATLYK-----EILTRAHVQEFSGVDDDHK 422
Db 1194 AFRTKVQDHSKAEDWKAAQVARGRQEAERKNLSLSSLEEEVSILNRQVLEKEGESKELKR 1253

QY 423 PIWMHAEREEMSKSRHHEGGTPTAEYGGWKACKVSSPTVNTLRLNLAGLYRRQKLEA 482
Db 1254 LVMAESEKSKLEES-----CACCRRQRPATVPPELQNAALLCGRRRCRAG 1298

QY 483 AETLEECALRSRRQGTDPISQTKVAELLGSDGRRTSQEGPGDSVKFEGGEDASVAVWS 542
Db 1299 REA-EKQRVASENLRLQELTSAERAEEELG-----QELKAWQEKFFQKEQALSTLQLE 1349

QY 543 GDGSGTL-----QSGSLGKIR-DVLRSSSEL--L 569
Db 1350 HTSTQALVSELLPAKHLCCQQLQAEQAAAEKRRHEELEQSKQAAGSLRAELLRQRELDEL 1409

QY 570 VRKLQGTGTPRPSSNMKRAASNLNQLNPSAAPLQVSRGLSA 610
Db 1410 IPLRQKVAEQDRTAQQLRAEKASYAEQLSM--LKKAHGLLA 1448

RESULT 5
US-08-483-924-4
; Sequence 4, Application US/08483924
; Patent No. 5882876
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-483-924-4

Query Match 5.3%; Score 167.5; DB 2; Length 2101;
Best Local Similarity 20.5%; Pred. No. 5.9e-06;
Matches 156; Conservative 106; Mismatches 272; Indels 227; Gaps 32;

QY 15 HRLSQEELGSTRLVSQGLEA-----LRSEHQAVLQSLSTI-----ECLQ--- 55
Db 750 HKRERKE-LEERAGRKGLARLLQLGEAHQAETEVLRRLEAEAMAAQHTAESECEQLVK 808

QY 56 -----QGGHEE-----GLVHEKARQLRRSMENIELGLSEAQVMLA-LASH-----LS 96
Db 809 EVAAWRDGYEDSQEEAQYGAMFQEQMLTLKEEKEKARQELQEAKEKVAGIESHSELQIS 868

QY 97 TVESEKQKLRAQVRRLCQE-----NQWLRDELQAGTQORLQSEQAQAOLE----- 141
Db 869 RQONKLAELHANLARALQQVQVEKVRQAQLADDLSTLQEKMAATSKAVARLETLVRKAGE 928

QY 142 -EEKKHLEFLGQ-LROYDEDGHTSEKEGD---ATKDSLDDLPFNEEEEDPSNGLSRGQG 196
Db 929 QQETASRELKVPARAGDRQPEWLEEQGGRQFCSTQAALQAM--ERAEQMGNELERLRA 986

QY 197 ATAAQQGGYEIPARLRTLHNLVQYAAQGRYEVAVPLCKQALDLELRTSGRHPDVATML 256
Db 987 ALMESQGG-----QOERGGQE-----REVARTQERGRAQADLAEK 1024

QY 257 NILA-LVYRDQNKYKE-----AAHLNDALSIRESTLGPDPHPAVAATLNNLAVLYG-K 307
Db 1025 AARAELEMLQNALNEQORVEFATLQELAHALTEKE---GKD-----QELAKLRGLE 1073

QY 308 RGKYKEAEPLCQRALEIREKV-----LGTNHPDVAK-----QLNNLA 344
Db 1074 AAQIKELEELRQTVKQLKEQLAKKEKEHSGSGAQSEAAGRTPTGPKLEALRAEVSKLE 1133

QY 345 LLCQNGKY-EAVERYIQ-----RALAIYEGQ-----LGPDPNPVARTKNNLA 386
Db 1134 QQCQKQEQADSLERSLEAERASRAERDSALETLQGLEKAQELGHSQSALASARELA 1193

QY 387 SCYLKQGYAEATLYK-----EILTRAHVQEFSGVDDDHK 422
Db 1194 AFRTKVQDHSKAEDWKAAQVARGRQEAERKNLSLSSLEEEVSILNRQVLEKEGESKELKR 1253

QY 423 PIWMHAEREEMSKSRHHEGGTPTAEYGGWKACKVSSPTVNTLRLNLAGLYRRQKLEA 482
Db 1254 LVMAESEKSKLEES-----CACCRRQRPATVPPELQNAALLCGRRRCRAG 1298

QY 483 AETLEECALRSRRQGTDPISQTKVAELLGSDGRRTSQEGPGDSVKFEGGEDASVAVWS 542
Db 1299 REA-EKQRVASENLRLQELTSAERAEEELG-----QELKAWQEKFFQKEQALSTLQLE 1349
```

QY 543 GDGSGTL-----QSGSLGKIR-DVLRSSSEL--L 569  
Db 1350 HTSTQALVSELLPAKHLCCQLQAEQAAAEKRRHREELQSKQAAGGLRAELRAQRELDEL 1409  
QY 570 VRKLGTEPRPSSNNMKRAASLNVLNQPSPAAPLOVSRGLSA 610  
Db 1410 IPLRQKVAEQRTAQQLRAEKASIAEQLSM--LKKAHGLLA 1448

RESULT 6  
PCT-US93-06160-4  
; sequence 4, Application PC/TUS9306160  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA HURWITZ & THIBEAULT  
; STREET: 53 STATE STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/06160  
; FILING DATE: 19930621  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/248-7000  
; TELEFAX: 617/248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US93-06160-4

Query Match 5.3%; Score 167.5; DB 4; Length 2101;  
Best Local Similarity 20.5%; Pred. No. 5.9e-06;  
Matches 156; Conservative 106; Mismatches 272; Indels 227; Gaps 32;  
QY 15 HRLSQEEILGSTRVLSQGLEA---LRSEHQAVLSLSQTI-----ECIQ--- 55  
Db 750 HRRERKE-LEERAGRGLEARLLQLGEAQAETEVLRRELAEMAAQAHTAESECEQLVK 808  
QY 56 -----QGGHEE-----GLVHEKARQLRRSMENIELGLSQAQVMLA-LASH-----LS 96  
Db 809 EVAARDGYEDSQEAEQYGAQFQELMTKEECEKARQELQEAKEKVAGIESHSELQIS 868  
QY 97 TVESEKQLRAQVRRLCQE-----NQWLRDELAGTQQLRSEQAQAQLE----- 141  
Db 869 RQONKLAELHANLARAQQVQVEKEVRAQKLADDLSTLQEKMAATSKAVARLETTLVRKAGE 928  
QY 142 -EEKKHLEFLGQ-LRQYDEDGHTSEEKGD---ATKDSLDDLPFNEEEDPSNGLSRGQG 196  
Db 929 QQETASRELKPARAGDRQPEWLEEQGQFCSTQAALQAM--EREAQMGNELERLRA 986  
QY 197 ATAAOQGGYEIPARLRLHNLNVIQYAAQGRYEVAVPLCKQALEDLERTSGRHPDVATML 256  
Db 987 ALMESQGG-----QQEERGQEE-----REVARLTQERGRAQADLALEK 1024

QY 257 NILA-LVYRDQNKYKE-----AAHLNDALSIRESTLGPDPHPAATLNNAVLYG-K 307  
Db 1025 AARAELEMLQNALNEQVREFATLQEAALAHALTEKE--GKD-----QELAKLRGLE 1073  
QY 308 RGKYKEAEPLCQRALEIREKV-----LGTNHPDVAK-----QLNNLA 344  
Db 1074 AAQIKELEELRQTVKQLKEQLAKKEKEHAGSGGAQSEAAAGRTPTGPKLEALRAEVSKLE 1133  
QY 345 LLCQNOQKY-EAVERYQ-----RALAIVEGQ-----LCPDNPVARTKNNLA 386  
Db 1134 QOCQKQEQEADSLERSLEAERASRAEDSALETQOQLEEKAEQELGHSQSALASAQRELA 1193  
QY 387 SCYLKQGYAEAEATLYK-----EILTRAHVQEEFGSVDDDDHK 422  
Db 1194 AFRTKVQDHKAEDWKAQVARGQEAERKNSLISSLEEEVSILNRQVLEKEGESKELKR 1253  
QY 423 PIWMAEEREEMSKSRHHEGTPYAEGGWYKACKVSSPTVNTTLNLGALYRRQKLEA 482  
Db 1254 LVMAESEKSKLEES-----CACCQRQPAVPELQNAALLCGRRCRASG 1298  
QY 483 AETLEECALPSRRQGTDPISQTKVAELLGESDGRRTSQEGPGDSVKFEGGEDASVAVEWS 542  
Db 1299 REA-EKQVASENLRLQELTSQAEAEELG-----QELKAWQEKFFQEQALSTLQLE 1349  
QY 543 GDGSGTL-----QSGSLGKIR-DVLRSSSEL--L 569  
Db 1350 HTSTQALVSELLPAKHLCCQLQAEQAAAEKRRHREELQSKQAAGGLRAELRAQRELDEL 1409  
QY 570 VRKLGTEPRPSSNNMKRAASLNVLNQPSPAAPLOVSRGLSA 610  
Db 1410 IPLRQKVAEQRTAQQLRAEKASIAEQLSM--LKKAHGLLA 1448

RESULT 7  
US-08-649-046-2  
; Sequence 2, Application US/08649046  
; Patent No. 5912415  
; GENERAL INFORMATION:  
; APPLICANT: OLSZEWSKI, NEIL E.  
; APPLICANT: JACOBSEN, STEVEN E.  
; TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF  
; TITLE OF INVENTION: IDENTIFICATION AND USE  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.  
; STREET: 119 NORTH FOURTH STREET, SUITE 203  
; CITY: MINNEAPOLIS  
; STATE: MINNESOTA  
; COUNTRY: USA  
; ZIP: 55401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/649,046  
; FILING DATE: 16-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MCCORMACK, MYRA H.  
; REGISTRATION NUMBER: 36,602  
; REFERENCE/DOCKET NUMBER: 110.00340101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-305-1225  
; TELEFAX: 612-305-1228  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 981 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein







Db 588 DQ-----LLQQLTELQSERQTSQVVTDRERLRLDRLQRIQYEGYETELARDVALCAA 639  
QY 308 ---RGYKEA---EPLCORALE---IREKVLGTNHPDVAKQLNNLALLCQNOQKYEAV 356  
Db 640 QEMEARYHAAVFLHQTLLLELATEWEDALRERALEAERDEAAAAELDAAASTSQN-ARESAC 698  
QY 357 ERYQRALAIYEGQLGPDNPVNARTKNNL-----ASCYLKQGYAEATLYKEIL-TRAH 410  
Db 699 ER-----LTSLEQQLRESEERAAELASQLEATAAAKSSAEQDRENTRATLEQQLRESEAR 753  
QY 411 VQEFSGVDDHDKPIWMHAEEREEMSKSRHHEGTPYAEYGGWKYKACKVSSPTVNTTLRNL 470  
Db 754 AAELASQLEATAAAKMSAEQDRENTRATLEQQLRDSEE-----RAAELASQLESTAAKM 808  
QY 471 GALYRRQKLEAAETLEECALRSRRQGTDPISQ---TKVAELLGESDGRRTS-----Q 520  
Db 809 SA---EQDRESTRATLEQQLRDSEERAAELASQLESTAAKMSAEQDRESTRATLEQQLR 865  
QY 521 EGPQDSVKFEGGEDASVAVEMSGDGSGLTQRSGLKIRDVLRSSSELLVRKLOGTEPRP 580  
Db 866 ESEERAAELASQLESTAAKMSAEQDRESTRATLEQQLRDSEERAAE-LASQLEATAAAK 924  
QY 581 SSSNMKRAASLNYLNQ 596  
Db 925 SSAEQDRENTRAALEQ 940

RESULT 11  
US-08-282-845-2  
; Sequence 2, Application US/08282845  
; Patent No. 5719263  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: A 230kd Antigen Present in Leishmania  
; TITLE OF INVENTION: Species  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple Macintosh Operating System 7.1  
; SOFTWARE: Microsoft Word for Macintosh 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/282,845  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/006,676  
; FILING DATE: JANUARY 15, 1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia Anne  
; REGISTRATION NUMBER: 34,693  
; REFERENCE/DOCKET NUMBER: 5004-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)587-0430  
; TELEFAX: (206)233-0644  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 955 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-282-845-2

Query Match 4.7%; Score 149.5; DB 1; Length 955;

Best Local Similarity 20.9%; Pred. No. 7.1e-05;  
Matches 129; Conservative 92; Mismatches 274; Indels 121; Gaps 25;  
QY 49 QTIECLOQGGHEGLVH-----EKARQLRRSMENIELGLS--EAQVMLALASHLS 96  
Db 378 ETLSTLRYASRARDIVNAQVNEEDPRARRIRIREEQMEDMRQAMAGDPAYVSELKKKLA 437  
QY 97 TVESEKQKLRQVRRRLCOE---NQWLRLDELQAGTQRLQRSEQAQVAAQLEEKHLEFLGQL 153  
Db 438 LLESEAQKRAADLQALEREREHNVQVERLLRLRATEAKSELESRAALQE-----EMTATR 492  
QY 154 ROYDE-DGHTSEEKGDATKDSLDDLFNPEEEEDPSNGLSRGQGATAAQGGYEIPARLR 212  
Db 493 ROADKMQALNRLKKEEQARKER--ELLKEMAKKDA--LSK-----VRRRKDAEIASERE 543  
QY 213 TLHNLVIQYAAQGR-YEVAVPL-----CKQALEDLERTSGRHGPDVATMLNIALVYR 264  
Db 544 KLESTVAQLEREQREVERALDQTHQRKQLQEALESSESTAAE-----R 587  
QY 265 DONKYKEAAHLLNDALSIRE-----STLGPDPHPAVAATLNNLAVLYGK----- 307  
Db 588 DQ-----LLQQLTELQSERQTSQVVTDRERLRLDRLQRIQYEGYETELARDVALCAA 639  
QY 308 ---RGYKEA---EPLCORALE---IREKVLGTNHPDVAKQLNNLALLCQNOQKYEAV 356  
Db 640 QEMEARYHAAVFLHQTLLLELATEWEDALRERALEAERDEAAAAELDAAASTSQN-ARESAC 698  
QY 357 ERYQRALAIYEGQLGPDNPVNARTKNNL-----ASCYLKQGYAEATLYKEIL-TRAH 410  
Db 699 ER-----LTSLEQQLRESEERAAELASQLEATAAAKSSAEQDRENTRATLEQQLRESEAR 753  
QY 411 VQEFSGVDDHDKPIWMHAEEREEMSKSRHHEGTPYAEYGGWKYKACKVSSPTVNTTLRNL 470  
Db 754 AAELASQLEATAAAKMSAEQDRENTRATLEQQLRDSEE-----RAAELASQLESTAAKM 808  
QY 471 GALYRRQKLEAAETLEECALRSRRQGTDPISQ---TKVAELLGESDGRRTS-----Q 520  
Db 809 SA---EQDRESTRATLEQQLRDSEERAAELASQLESTAAKMSAEQDRESTRATLEQQLR 865  
QY 521 EGPQDSVKFEGGEDASVAVEMSGDGSGLTQRSGLKIRDVLRSSSELLVRKLOGTEPRP 580  
Db 866 ESEERAAELASQLESTAAKMSAEQDRESTRATLEQQLRDSEERAAE-LASQLEATAAAK 924  
QY 581 SSSNMKRAASLNYLNQ 596  
Db 925 SSAEQDRENTRAALEQ 940

RESULT 12  
US-08-428-414A-3  
; Sequence 3, Application US/08428414A  
; Patent No. 5912166  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TITLE OF INVENTION: LEISHMANIASIS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/428,414A  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:  
NAME: Kadlec, Ann T.  
REGISTRATION NUMBER: 39,244  
REFERENCE/DOCKET NUMBER: 210121.407  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
TELEX: 3723836 SEEDANDBERRY  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 955 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-428-414A-3

Query Match 4.7%; Score 149.5; DB 2; Length 955;  
Best Local Similarity 21.1%; Pred. No. 7.1e-05;  
Matches 130; Conservative 91; Mismatches 274; Indels 121; Gaps 25;

QY 49 QTIECLOQGGHEGLVH-----EKARQLRRSMENIELGLS--EAQVMLALASHLS 96  
Db 378 ETLSTLYASRARDIVNVAQVNEPRARRIRELEEQMEDMRQAMAGDPAYVSELKKLA 437  
QY 97 TVESEKQKLRAQVRLCQE---NQWLRLDELACTQRLQRSEQAQAQLEEKKHLEFLGQL 153  
Db 438 LLESEAQKRAADLQALERERHNVQVRLRATAEAKSELESRAAALQE-----EMTATR 492  
QY 154 ROYDE-DGHTSEEKGDATKDSLDDLPNEEEEDPSNGLSRGOGATAAQGGYEIPARLR 212  
Db 493 ROADKMQALNRLKKEEQARKER--ELLKEMAKKDAA--LSKVRRLDA-----EIASERE 543  
QY 213 TLHNLVIQYAAQGR-YEVAVPL-----CKQALEDLERTSGRHPDPVATMLNILALVYR 264  
Db 544 KLESTVAQLEREQREVALDALQTHQRKLOEALLESERTAAE-----R 587  
QY 265 DQNYKEAAHLLNDALSIRE-----STLGPDPHPAVALNNLAVLYGK----- 307  
Db 588 DQ-----LLQQLTELQSEREQVVDRLTRDLQRIQYEGETELARDVALCAA 639  
QY 308 ---RGKYKEA---EPLCQRALE---IREKVLGTHNPDPVAKQLNNLALLCQNGKYEAV 356  
Db 640 QEMEARYHAAVFHLQTLLELATEWEDALRERALEAERDEAAAAELDAAASTSQN-ARESAC 698  
QY 357 ERYQRALAIYEGQLGPDNPNVARTKNNL-----ASCYLKQGYAEATLYKEIL-TRAH 410  
Db 699 ER-----LTSLEQQLRESEERAAELASQLEATAAAKSSAEQDRENTRATLEQQLRESEAR 753  
QY 411 VQFEGSVDDHDKPIWMHAEEREEMSKSRHHHEGTPYAEYGGWKYKACKVSSPTVTTLRNL 470  
Db 754 AAELASQLEATAAAKMSAEQDRENTRATLEQQLRDSE-----RAAELASQLESTTAAKM 808  
QY 471 GALYRRQCKLEAAETLEECALRSRRQGTDPISQ---TKVAELGEGSDGRTS-----Q 520  
Db 809 SA---EQDRESTRATLEQQLRDSEERAAELASQLESTTAAKMSAEQDRESTRATLEQQLR 865  
QY 521 EGPDSVKFEGGEDASVAVWEGSGGTIQRSGSLGKIRDVLRRSSELLVRKLGQTEPRP 580  
Db 866 ESEERAAELASQLESTTAAKMSAEQDRESTRATLEQQLRDSEERAAE-LASQLEATAAAK 924  
QY 581 SSSNMKRAASLNLYNQ 596  
Db 925 SSAEQDRENTRAALEQ 940

RESULT 13  
PCT-US94-00324-1  
Sequence 1, Application PC/TUS9400324  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven  
TITLE OF INVENTION: Diagnosis of Leishmaniasis  
NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word, version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/00324  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/006,676  
FILING DATE: 15-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia Anne  
REGISTRATION NUMBER: 34,693  
REFERENCE/DOCKET NUMBER: 5004-WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 955 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-00324-1

Query Match 4.7%; Score 149.5; DB 4; Length 955;  
Best Local Similarity 20.9%; Pred. No. 7.1e-05;  
Matches 129; Conservative 92; Mismatches 274; Indels 121; Gaps 25;

QY 49 QTIECLOQGGHEGLVH-----EKARQLRRSMENIELGLS--EAQVMLALASHLS 96  
Db 378 ETLSTLYASRARDIVNVAQVNEPRARRIRELEEQMEDMRQAMAGDPAYVSELKKLA 437  
QY 97 TVESEKQKLRAQVRLCQE---NQWLRLDELACTQRLQRSEQAQAQLEEKKHLEFLGQL 153  
Db 438 LLESEAQKRAADLQALERERHNVQVRLRATAEAKSELESRAAALQE-----EMTATR 492  
QY 154 ROYDE-DGHTSEEKGDATKDSLDDLPNEEEEDPSNGLSRGOGATAAQGGYEIPARLR 212  
Db 493 ROADKMQALNRLKKEEQARKER--ELLKEMAKKDAA--LSK-----VRRKDAEIASERE 543  
QY 213 TLHNLVIQYAAQGR-YEVAVPL-----CKQALEDLERTSGRHPDPVATMLNILALVYR 264  
Db 544 KLESTVAQLEREQREVALDALQTHQRKLOEALLESERTAAE-----R 587  
QY 265 DQNYKEAAHLLNDALSIRE-----STLGPDPHPAVALNNLAVLYGK----- 307  
Db 588 DQ-----LLQQLTELQSEREQVVDRLTRDLQRIQYEGETELARDVALCAA 639  
QY 308 ---RGKYKEA---EPLCQRALE---IREKVLGTHNPDPVAKQLNNLALLCQNGKYEAV 356  
Db 640 QEMEARYHAAVFHLQTLLELATEWEDALRERALEAERDEAAAAELDAAASTSQN-ARESAC 698  
QY 357 ERYQRALAIYEGQLGPDNPNVARTKNNL-----ASCYLKQGYAEATLYKEIL-TRAH 410  
Db 699 ER-----LTSLEQQLRESEERAAELASQLEATAAAKSSAEQDRENTRATLEQQLRESEAR 753  
QY 411 VQFEGSVDDHDKPIWMHAEEREEMSKSRHHHEGTPYAEYGGWKYKACKVSSPTVTTLRNL 470  
Db 754 AAELASQLEATAAAKMSAEQDRENTRATLEQQLRDSE-----RAAELASQLESTTAAKM 808  
QY 471 GALYRRQCKLEAAETLEECALRSRRQGTDPISQ---TKVAELGEGSDGRTS-----Q 520

Db 809 SA---EQDRESTRATLEQQLRDSEERAAELASQLESTTAAKMSAEQDRESTRATLEQQLR 865

QY 521 EGPQDSVKEGGEDASVAVESGSGTLQSGSLKIRDLVRRSELVVRKLGQTEPRP 580

Db 866 ESEERAAELASQLESTTAAKMSAEQDRESTRATLEQQLRDSEERAAE-LASQLEATAAK 924

QY 581 SSSNMKRAASLYNLQ 596

Db 925 SSAEQDRENTRAALEQ 940

RESULT 14

US-07-853-913-2

; Sequence 2, Application US/07853913

; Patent No. 5338839

; GENERAL INFORMATION:

; APPLICANT: McKay, Ronald D.G.

; APPLICANT: Lendahl, Urban

; TITLE OF INVENTION: Nestin Expression As An Indicator of

; TITLE OF INVENTION: Neuroepithelial Tumors

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

; STREET: Two Militia Drive

; CITY: Lexington

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02173

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/853,913

; FILING DATE: 19920319

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/660,412

; FILING DATE: 22-FEB-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/603,803

; FILING DATE: 25-OCT-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/201,762

; FILING DATE: 02-JUN-1988

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/180,548

; ATTORNEY/AGENT INFORMATION:

; NAME: Granahan, Patricia

; REGISTRATION NUMBER: 32,227

; REFERENCE/DOCKET NUMBER: MIT-4641AAAA

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-861-6240

; TELEFAX: 617-861-9540

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1805 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-853-913-2

Query Match 4.7%; Score 149.5; DB 1; Length 1805;

Best Local Similarity 20.8%; Pred. No. 0.0002;

Matches 165; Conservative 100; Mismatches 290; Indels 239; Gaps 37;

QY 6 LGQRDEPAGHRLSQE--EILGSTRLVSQGLEALRS-----EHQAVLQSLQSOTIECLQQG 57

Db 951 LEKESQDSGKSLEDESQETFGP--LEKENAESLRLAGQDQEEQKLEQETQQTILRAV--- 1005

QY 58 GHEEGLVH--EKA-----RQLRRSM--ENIELGLSEAQVMLALASHLST----- 97

Db 1006 GNEQMAVSPPEKVDPELPKPLGNDQEIARSLGKENQESLVSLEKGIETVKSLETEIIEP 1065

QY 98 VESEKQLRAQVRRRLCQENQWL-----RDELAGTQQLRQSEQAQAQLEEEKKH 146

Db 1066 LETAEEDLERRKSIDTQEPLNSTEVARETVPEPDEPGSLGSDENRETITLSLEKESQE 1125

QY 147 LEFLG-----QLRQYDEDDGHTSEEKQ-DATKDSLDDLPNEEEEDPSNGLSR 193

Db 1126 LSSLGKWNVETRVEDSQCLQVEEGLQEEHQESLREVKQELPSSGNQQRWEDVVEGKAV 1185

QY 194 GQGATAA-----QGGYEIIPARLRTLHNLVQYAAQGR---YEVAVPLC 234

Db 1186 GQEAPLATTGVTGTEDKAELHLRGQGGEEAAAAGELLQDIVGEAWSLGSSEPKQRPV-- 1243

QY 235 KQALEDLERTISGRGHPDVATMLNIALVYRDQNKY-----EAAHLNDALS 281

Db 1244 AEALDNLE---GGALEVPVAQSMPEVTERDEDRQAAGEQDSIEVTGLGEEA---RTGLE 1296

QY 282 IRESTLGPDHP-----AVAATLNNLAV-----LYGKRKYKEAEPLCQRALEIREK 327

Db 1297 LEQEVVGLDEPRHFAREEAIPPSLGESVAKIAQGLEPGKPKKEAGALDSGILEL--- 1353

QY 328 VLGTHNPDVAKQLNNLALLCQNGQKVEAVERYYQALAI---YEGQLGPDNPVARTKN 383

Db 1354 -----PKTSSE---ALECQGHESSEMEGWEEEAASLETSDHEGSDAP-QPRPPPETEE 1402

QY 384 NLASCYLKQGYABAEATLYK-----EILTRAH-----VQFEGSVDDDDHKPIWMHAE 430

Db 1403 DEGA---QAALTAPGPKLLEPCSPILITDAHELQPOAEGIQEAGWQPEAGSEALERVEN 1459

QY 431 REEMSKSRHHEG-----GTPYAEYGGWKACKVS----- 459

Db 1460 EPEFGLGEIPEGLQDWEEGRESEADDLGETLPDSTPL---GLYLRSPASPKWDLAGEQ 1515

QY 460 --SPTVNTTLRLNG-ALYRRQKLEAAETLEECALRSRRQGT----- 499

Db 1516 RLSPQGDAGKEDWGPVPAQAAGLSGPPEEEEE---QGHGSDLSEEFEDLGTEASLLP 1570

QY 500 -----PISQTKVAELLGESDG-----RRTSQEGPGDSVKFEGGEDASVAVEW 541

Db 1571 GVPKEVADHVGVQVPPVLQPCWDQGGESDGFADDEESGEEGEEEDADEEGAESGA---QW 1627

QY 542 SGDGSGLTQSGSLGKIRDLVRRSEL---VRKLQGTETPRPSSNNMKRAASLYNLNQ 597

Db 1628 WSGA-----SGGGCKVDIAQRGDPVQESVSGVSLWDDGLRGAANV---PALEMVSQD 1679

QY 598 SAAPLOVSRGLSAS 611

Db 1680 SAEPGSEESAS 1693

RESULT 15

US-08-328-254-6

; Sequence 6, Application US/08328254

; Patent No. 5710022

; GENERAL INFORMATION:

; APPLICANT: Zhu, Xueliang

; APPLICANT: Lee, Wen-Hwa

; TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25



Search completed: August 15, 2000, 10:53:34  
Job time: 23119 sec

;  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,254  
; FILING DATE: 24-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/141,239  
; FILING DATE: 22-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-CJ 1191  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2482 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-328-254-6

Query Match 4.7%; Score 148; DB 1; Length 2482;  
Best Local Similarity 21.2%; Pred. No. 0.00047;  
Matches 137; Conservative 112; Mismatches 256; Indels 140; Gaps 30;

QY 16 RLS--QEEILGSTRVSGLEALRSEHQAVLSQLSQTIECLQQGGHEEGLVHEKA---R 69  
Db 1474 RLSSTQEEV---HQLRRGIEKLRVRIEADKKQLHIAEKLKERERENDSLKDKVENLER 1529  
QY 70 QLRRSMENIELGL-----SEAQV-----MLALASHLSTVESEKQKLRQAQVRR 111  
Db 1530 ELQMSSEENQELVILDAENSKAEVETLKTQIEEMARSLKVFELDLVTLRSEKENLTQI-- 1587  
QY 112 LCQENQWLDE---LAGTQQRQLORSEQAVAQLEEEKK-HLEFL-GQLROYDE--DGHTS 163  
Db 1588 --QEQQQLSELDKLSSFKSLLEEKEQAEIQIKESKTAVEMLQNLQKELNEAVAALCG 1645  
QY 164 EEKEGDATKDSLDDLPFNEEEEDPSNGLSRGGATAAQGGYEIPARLRTLHN---LVIQ 220  
Db 1646 DQEIMKATEQSLDP--PIEEHQLRNSIEK-----LRARLEADEKKQLCVLQ 1690  
QY 221 YAAQGRYEVAVPLCKQALEDLER--TSGRGHPDVATMLNIALVYRDQNKYKEAAHLLND 278  
Db 1691 QKSESEHH--ADLLKGRVENLERELEIARTNQEHAAAL-----EAENSKGE 1733  
QY 279 ALSIRESTLGPDPHPAVAATLNNLAVLYKRGKYKEABPLCQRALEIREKVLGTNHPDDVAK 338  
Db 1734 VETLKAKIEGWTQSLRGLELDVVTIRSEKENLTNELQKEQERISEL--EIINSSFENILQ 1791  
QY 339 QLNNLALLCQNOQGYEAVERYYQRALIYEGQLGPDNPVARTKNNLASCYLK-QGKYAE 397  
Db 1792 E-----KEQEKVQMKKEK-SSTAMEMLTQLKELNERVAALHNDQEACKAKEQNLSQ 1842  
QY 398 AETLYKE-----ILTRAHV-----QEFGSVDDDHKPIWMHAEER 431  
Db 1843 VECLELEKAQLLQGLDEAKNNYIVLQSSVNGLIQEVEDGKQLEKKDEEISRLKNQIQDQ 1902  
QY 432 EEM-SKSRHHEGGTPYAEYGGWYKACKVSSPTVNTTLRNLGALYRR-----QCKLEAAET 485  
Db 1903 EQLVSKLSQVEG-----EHQLWKEQ-----NLELRNLTVLEQKIQVLOSKNASLQD 1949  
QY 486 LEECALRSRRQGTDPISQTKVAEL-LGESDGRRTSQEGPGDSVKFEGGEDASVAVENWSGD 544  
Db 1950 TLEVQSSYKYNLENELELTKMDKMSFVEKYNKMTAKE---TELQREMHMAQKTAELQEE 2006  
QY 545 GSGTLQR-SGSLGKTRDVLRRSSELLVR-KLQGTETPRPSSSNMKR 587  
Db 2007 LSGEKNRLAGELQLLLEEIKSSKDKQLKELTLENSELKKSILDCMHK 2051



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2000, 04:31:13 ; Search time 131.1 Seconds  
(without alignments)  
292.194 Million cell updates/sec

Title: US-09-036-614A-1  
Perfect score: 3161  
Sequence: 1 MSGLVLGQRDEPAGHRLSQE.....APLQVSRGLSASTMDLSSSS 619

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2000	63.3	569	I53013	kinesin light chain
2	1985.5	62.8	560	C41539	kinesin light chain
3	1956	61.9	551	B41539	kinesin light chain
4	1954.5	61.8	542	A41539	kinesin light chain
5	1822	57.6	686	S33815	kinesin light chain
6	1821	57.6	686	S36727	kinesin light chain
7	1809.5	57.2	649	S33813	kinesin light chain
8	1803.5	57.1	677	S33814	kinesin light chain
9	1520	48.1	537	T34113	hypothetical prote
10	1517	48.0	563	S47997	kinesin light chain
11	1516	48.0	540	S41864	kinesin light chain
12	1491.5	47.2	522	S41865	kinesin light chain
13	1453	46.0	451	S33816	kinesin light chain
14	965	30.5	516	T23827	hypothetical prote
15	241.5	7.6	114	S47998	kinesin light chain
16	232	7.3	609	T01892	hypothetical prote
17	229.5	7.3	886	T35469	probable ATP /GTP-
18	189	6.0	342	E70463	conserved hypothet
19	189	6.0	948	S75991	hypothetical prote
20	185	5.9	1791	T02909	hypothetical prote
21	181	5.7	1939	T18372	repeat organellar
22	179.5	5.7	2176	T13806	toucan gene protei
23	179	5.7	1963	T20770	hypothetical prote
24	179	5.7	1966	1 MWKW	myosin heavy chain
25	175.5	5.6	671	A40692	signal recognition
26	173.5	5.5	848	A44972	paramyosin - nemat
27	167.5	5.3	2101	A42184	nuclear mitotic ap
28	167.5	5.3	3187	JC5837	364K Golgi complex
29	166.5	5.3	879	A48575	paramyosin - nemat

30	166	5.3	1938	2	JC5421	smooth muscle myos
31	166	5.3	1972	1	A41604	myosin heavy chain
32	166	5.3	1972	2	JC5420	smooth muscle myos
33	163	5.2	2168	2	T30171	ninein - mouse
34	162.5	5.1	1300	2	T18364	ro-3 protein - Neu
35	162	5.1	1969	1	S02771	myosin heavy chain
36	162	5.1	1992	2	T23622	hypothetical prote
37	162	5.1	2442	2	T08621	centrosome associa
38	161.5	5.1	1133	2	T22976	hypothetical prote
39	160.5	5.1	1426	2	T00337	hypothetical prote
40	160	5.1	1920	2	A53188	pericentrin - mous
41	159.5	5.0	543	2	A38093	transformation-sen
42	159	5.0	1974	2	T30010	hypothetical prote
43	157	5.0	886	2	H69378	purine NTPase homo
44	157	5.0	3225	2	I52300	giantin - human
45	156.5	5.0	1390	2	T14004	trfA protein - sli

ALIGNMENTS

RESULT 1

I53013

kinesin light chain - human

C:Species: Homo sapiens (man)

C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999

C:Accession: I53013

R:Cabeza-Arvelaiz, Y.; Shih, L.C.; Hardman, N.; Asselbergs, F.; Bilbe, G.; Schmitz, A

DNA Cell Biol. 12, 881-892, 1993

A:Title: Cloning and genetic characterization of the human kinesin light-chain (KLC)

A:Reference number: I53013; MUID:94099888

A:Accession: I53013

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-569 <RES>

A:Cross-references: GB:L04733; NID:g307084; PIDN:AAAL6576.1; PID:g307085

Query Match 63.3%; Score 2000; DB 2; Length 569;  
Best Local Similarity 69.8%; Pred. No. 9.1e-109;  
Matches 395; Conservative 77; Mismatches 76; Indels 18; Gaps 8;

QY 1 MSGVLVQGRDEPAGHRLSQEILGSTRVLSQGLEALRSEHQAVLQSLSQTIETCLQQGGHE 60

Db 1 MSTVMYIKEDKL--EKLTDDEIISKTKQVIOGLEALKNEHNSILQSLLETCLKKK--DDE 57

QY 61 EGLVHEKARQLRRSMENIELGLSEAQVMLALASHLSTVESEKQKLRQAVRRLCQENQWLR 120

Db 58 SNLVEEKSNMIRKSLEMLELGLSEAQVMMALSNHNLNAVESEKQKLRQAVRRLCQENQWLR 117

QY 121 DELAGTQORLQRSEQVAQLEEEKKHLEFLGQLROYDEDEGHTSEEKGDATKDSLDDLFP 180

Db 118 DELANTQOKLQKSEQSVAQLEEEKKHLEFMNLKYYDDDISPEDKDTDTKEPLDDLFP 177

QY 181 NEEEDPSNGLSR--GQATAAQGGYEIPARLRLTNLNLVIQYAAQGRYEVAVPLCKQAL 238

Db 178 N-DEDDPGQGIQQHSSAAAAAQGGYEIPARLRLTNLNLVIQYASQGRYEVAVPLCKQAL 236

QY 239 EDLERTSGRHPDVATMNLALVYRDQNKYKFAAHLNDALSIRESTLGPDPHVAATL 298

Db 237 EDLEKTSRHDHPDVATMNLALVYRDQNKYKFAAHLNDALSIREKILGKDHVAATL 296

QY 299 NNLAIVLYGKRGKYKEAFPLCQRALEIREKVLGTNHPDVAKQLNNLALCQNGKYEAVE 358

Db 297 NNLAIVLYGKRGKYKEAFPLCQRALEIREKVLGKDHDPDVAKQLNNLALCQNGKYEVEY 356

QY 359 YYQRALAIYEGQLGPDNPVARTKNNLASCVLKQGYAEAEETLYKEILTRAHVQEGSVD 418

Db 357 YYQRALEIYQTKLGPDDPNVAKTKNNLASCVLKQGYKQAEETLYKEILTRAHEREFGSVD 416

QY 419 DDHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWYKACKVSSPTVNTTLNLGALYRRQ 478

Db 417 DENKPIWMHAEERECKGKQ--KDGTSFGYGGWYKACKVDSPTVTTLNLGALYRRQ 474







Db 121 WLRDELAATQKLOTSEQNADLEVYKHKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177

QY 178 LFPNEEEEDPSNGLS-----RQGATAAQQGGYEIPARLRTLHNLVVIQYAAQGRYEV 229

Db 178 GFP-----EDDDGGQADESYPPQOTGSGSVSAAAGGYEIPARLRTLHNLVVIQYASQRYEV 233

QY 230 AVPLCKQALEDLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLNDALSIRESTLGP 289

Db 234 AVPLCKQALEDLERTSGHDHPDVATMLNIALVYRDQNKYKEAGNLHLDAIRAIREKTLP 293

QY 290 DHPAVALTNLNLAVLYGKRGKYKEAEPICQRALEIREKVLGTHNHPDVAKQLNNLALLCQN 349

Db 294 DHPAVALTNLNLAVLYGKRGKYKEAEPICQRALEIREKVLGTHNHPDVAKQLNNLALLCQN 353

QY 350 QGKYEAVERYQORALAIYEGQLGPDNPNVARTKNNLASCYLKQGYAEATLYKEILTRA 409

Db 354 QGKYEEVEWYQORALEIYEKKGPDNPNVARTKNNLAAAAYLKQGYKAAETLYKQVLTRA 413

QY 410 HVQEFQ-SVDD-DHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWYKACKV-----SSPT 462

Db 414 HEREFGLSADDDKDKNKPIMQAEEREKGFKN---APYGDYGGWHKAACKVDSRSRSPT 470

QY 463 VNTTLNLGALYRRQKLEAAETLEECALRSRRQGTDPISQTKVAELLG-----511

Db 471 VTTTLKNLALYRRQKGYDAEILEECAMKSRNALDMVRETKVRELLGQDLSTDVPRSE 530

QY 512 -----ESDGRRTS---QEGPGDSVKFE---GGEDASVAVVWSG-----543

Db 531 AMAKERHRRSSGTPRHGSTESVSYEKTGSEEVSGVAVKAKKAKDRSIPAGYVEI 590

QY 544 -----DGSGLTQRLSGSLGKIRDLVRRSSSELLVRKLGQTEPRPSSSNMKRAASLNY 593

Db 591 PRSPPHVLVENGDKLRRSGSLSKLRASVRRSSTKLLNKLKGRE-SDDGGMKRASSMSV 649

QY 594 L-----NQPSAAPLQVS-RGLSASTMDLSS 617

Db 650 LPSRGNDSTPAPIQLSQRGVGRVSHDNLSS 679

RESULT 7

S33813

kinesin light chain - sea urchin (Strongylocentrotus purpuratus)

C:Species: Strongylocentrotus purpuratus (purple urchin)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997

C:Accession: S33813

R:Wedaman, K.P.; Knight, A.E.; Kendrick-Jones, J.; Scholey, J.M.

J. Mol. Biol. 231, 155-158, 1993

A:Title: Sequences of sea urchin kinesin light chain isoforms.

A:Reference number: S33813; MUID:93267648

A:Accession: S33813

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-649 <WED>

A:Cross-references: EMBL:L10233; NID:g161525; PID:g161526

Query Match 57.2%; Score 1809.5; DB 2; Length 649;

Best Local Similarity 59.2%; Pred. No. 1.2e-97;

Matches 392; Conservative 81; Mismatches 124; Indels 65; Gaps 16;

QY 1 MSGVLVQGRDEPAGHR--LSQEEILGSTRLVSQGLEALRSEHQAVLQSLQTIETCLQGG- 57

Db 1 MSGSKLSTPNNSGGQGNLSQEQITGTREVIGKLEQLKNEHNDILNSLYQSLKMLKKDT 60

QY 58 GHEEGLVHEKARQLRRSMENIELGLSEAQVLMALASHLSTVESEKQKLAQVRRLLCQENQ 117

Db 61 PGDSNLVEEKTDIIEKSLESLELGLGEAKVMALGHLNMVVEAEKQKLAQVRRLLVQENT 120

QY 118 WLRDELACTQORLQORSEQAQVAOLEEKKHLEFLGQLROYDEGHTSEEKGDATKDSLDD 177

Db 121 WLRDELAATQKLOTSEQNADLEVYKHKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177

QY 178 LFPNEEEEDPSNGLS-----RQGATAAQQGGYEIPARLRTLHNLVVIQYAAQGRYEV 229

Db 178 GFP-----EDDDGGQADESYPPQOTGSGSVSAAAGGYEIPARLRTLHNLVVIQYASQRYEV 233

QY 230 AVPLCKQALEDLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLNDALSIRESTLGP 289

Db 234 AVPLCKQALEDLERTSGHDHPDVATMLNIALVYRDQNKYKEAGNLHLDAIRAIREKTLP 293

QY 290 DHPAVALTNLNLAVLYGKRGKYKEAEPICQRALEIREKVLGTHNHPDVAKQLNNLALLCQN 349

Db 294 DHPAVALTNLNLAVLYGKRGKYKEAEPICQRALEIREKVLGTHNHPDVAKQLNNLALLCQN 353

QY 350 QGKYEAVERYQORALAIYEGQLGPDNPNVARTKNNLASCYLKQGYAEATLYKEILTRA 409

Db 354 QGKYEEVEWYQORALEIYEKKGPDNPNVARTKNNLAAAAYLKQGYKAAETLYKQVLTRA 413

QY 410 HVQEFQ-SVDD-DHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWYKACKV-----SSPT 462

Db 414 HEREFGLSADDDKDKNKPIMQAEEREKGFKN---APYGDYGGWHKAACKVDSRSRSPT 470

QY 463 VNTTLNLGALYRRQKLEAAETLEECALRSRRQGTDPISQTKVAELLG-----511

Db 471 VTTTLKNLALYRRQKGYDAEILEECAMKSRNALDMVRETKVRELLGQDLSTDVPRSE 530

QY 512 -----ESDGRRTS---QEGPGDSVKFE---GGEDASVAVVWSG-----561

Db 531 AMAKERHRRSSGTPRHGSTESVSYEKTGSEEVSGVAVKAKKAKDRSIPAGYVEI 581

QY 562 LRRSELLVRKLGQTEPRPSSSNMKRAASLNYL-----NQPSAAPLQVS-RGLSASTMDL 615

Db 582 VRRSSTKLLNKLKGRE-SDDGGMKRASSMSVLPSSRGNDSTPAPIQLSQRGVGRVSHDNL 640

QY 616 SS 617

Db 641 SS 642

RESULT 8

S33814

kinesin light chain - sea urchin (Strongylocentrotus purpuratus)

C:Species: Strongylocentrotus purpuratus (purple urchin)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997

C:Accession: S33814

R:Wedaman, K.P.; Knight, A.E.; Kendrick-Jones, J.; Scholey, J.M.

J. Mol. Biol. 231, 155-158, 1993

A:Title: Sequences of sea urchin kinesin light chain isoforms.

A:Reference number: S33813; MUID:93267648

A:Accession: S33814

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-677 <WED>

A:Cross-references: EMBL:L10234; NID:g161527; PID:g161528

Query Match 57.1%; Score 1803.5; DB 2; Length 677;

Best Local Similarity 57.9%; Pred. No. 2.8e-97;

Matches 394; Conservative 82; Mismatches 130; Indels 75; Gaps 16;

QY 1 MSGVLVQGRDEPAGHR--LSQEEILGSTRLVSQGLEALRSEHQAVLQSLQTIETCLQGG- 57

Db 1 MSGSKLSTPNNSGGQGNLSQEQITGTREVIGKLEQLKNEHNDILNSLYQSLKMLKKDT 60

QY 58 GHEEGLVHEKARQLRRSMENIELGLSEAQVLMALASHLSTVESEKQKLAQVRRLLCQENQ 117

Db 61 PGDSNLVEEKTDIIEKSLESLELGLGEAKVMALGHLNMVVEAEKQKLAQVRRLLVQENT 120

QY 118 WLRDELACTQORLQORSEQAQVAOLEEKKHLEFLGQLROYDEGHTSEEKGDATKDSLDD 177

Db 121 WLRDELAATQKLOTSEQNADLEVYKHKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177

QY 178 LFPNEEEEDPSNGLS-----RQGATAAQQGGYEIPARLRTLHNLVVIQYAAQGRYEV 229

Db 178 GFP-----EDDDGGQADESYPPQOTGSGSVSAAAGGYEIPARLRTLHNLVVIQYASQRYEV 233

QY 230 AVPLCKQALEDLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGP 289  
D 234 AVPLCKQALEDLERTSGRHPDVATMLNIALVYRDQNKYKEAGNLLHDALAIKRETLGP 293  
QY 290 DHPAVALTNLNLAVLYGKRGKYKEAEPICQRALEIREKVLGTNHPDVAKQLNNLALCQN 349  
D 294 DHPAVALTNLNLAVLYGKRGKYKEAEPICQRALEIREKVLGTNHPDVAKQLNNLALCQN 353  
QY 350 QGYEAEVYQYRALIYEGQLGPDNPVARTKNNLASCYLKQGYAEATLYKEILTRA 409  
D 354 QGYEAEVYQYRALIYEGQLGPDNPVARTKNNLAAAYLKQGYAAETLYKQVLTTRA 413  
QY 410 HVQFEG-SVDD-DHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWKYKACKV-SSPT 462  
D 414 HERFGLSADDDKPKIWMHAEEREEMSKSRHHEGGTPYAEYGGWKYKACKV-SSPT 470  
QY 463 VNTTLNLGALYRRQGLKLEAAETLEECALRSRQGTDPISQTKVAELLG----- 511  
D 471 VTTTLNLGALYRRQGLKLEAAETLEECALRSRQGTDPISQTKVAELLG----- 530  
QY 512 ----ESDGRRTS---QEGPGDSVKFE---GGEDASVAVESW----- 542  
D 531 AMAKRHRHRRSSGTPRHGSTEVSVEYKTDGSEKRAKADRSRIPAGYVEIPRSPHVLV 590  
QY 543 GDGSGTLQSGSLKIRDLVLRSSSELLVRLQGTPEPRSSNNMKRAASLNYL-----NQP 597  
D 591 ENGDKLRRSGSLKIRASVRRSSSTKLLNKLKGRE-SDDGGMKRASSMSVLPGRGNDES 649  
QY 598 SAAPLQVS-RGLSASTMDLSS 617  
D 650 TPAPQLSQGRVGRVSHDNLSS 670

RESULT 9  
T34113  
hypoetical protein C18C4.10 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34113  
R:Gattung, S.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid C18C4.  
A:Reference number: 221478  
A:Accession: T34113  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-537 <GAT>  
A:Cross-references: EMBL:U55369; PIDN:AAC25830.1; GSPDB:GN00023; CESP:C18C4.10  
A:Experimental source: strain Bristol N2; clone C18C4  
C:Genetics:  
A:Gene: CESP:C18C4.10  
A:Map position: 5  
A:Introns: 37/1; 142/3; 449/2; 485/3

Query Match 48.1%; Score 1520; DB 2; Length 537;  
Best Local Similarity 58.4%; Pred. No. 5.7e-81;  
Matches 313; Conservative 75; Mismatches 106; Indels 42; Gaps 9;

QY 17 LSQEEILGSTRVLSQGLEALRSEHQAVLQSLQTECLQGGHVEGHEKARQLRRSME 76  
D 1 MSQDDVTGLRTVQOGLAEALREHSTISNTLETSTVKGVE--DEAPLPKQLSQINDNLD 58  
QY 77 NIELGLSEAQVMLALASHLSTVESEKQLRAQVRLCQENQWLRLDELACTQORLORSEQA 136  
D 59 KLVCVGVDETSLMLVFQLTQGMADAHQYQARRRRCQENAWLRDELSTQIKLQSQEQM 118  
QY 137 VAQLEEKHLEFLGQLROYDEDDGHTSEEK-----DATKDSLDDL-FPNEEEED--- 186  
D 119 VAQLEENKHLKYMASIKQFD-DGTQSDTKTSVDVGPQVPTNETLQELGFGPEDEEDMNA 177  
QY 187 -----PSNGLSRGQGATAAQGGYEIPARLRTLNLVLIQYAAQGRYEVAVPLCKQALE 239

D 178 SQFNQPTPANQMA-----ASANVGYEIPARLRTLNLVLIQYASQGRYEVAVPLCKQALE 231  
QY 240 DLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPAVALTN 299  
D 232 DLEKTSRHDHPDVATMLNIALVYRDQNKYKEAANLLNEALSIREKCLGESHHPAVALTN 291  
QY 300 NLAVLYGKRGKYKEAEPICQRALEIREKVLGTNHPDVAKQLNNLALCQNQGYEAEVY 359  
D 292 NLAVLYGKRGKYKEAEPICQRALEIREKVLGTNHPDVAKQLNNLALCQNQGYEAEVY 351  
QY 360 YQALAIYEGQLGPDNPVARTKNNLASCYLKQGYAEATLYKEILTRAHVQFEGSVDD 419  
D 352 YKRALEIYESKLGPDNPVARTKNNLSSAYLKQGYKEAEELYKQILTRAHEREFGQISG 411  
QY 420 DHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWKYKACKVSSPTVNTTLNLGALYRRQK 479  
D 412 ENKPIWQIAEEREE---NKHKGEGATANEQAGWAKAAKVDSPVTTLNLGALYRRQK 468  
QY 480 LEAAETLEECALRSRQ-----GTDPISTQTKVAELLGESDG---RRTSQEG 522  
D 469 YEAETLEDVALRAKKQHEPLRSRGAMGIDEMSQMMASTIGGSRNSMTSTSTQIG 524

RESULT 10  
S47997  
kinesin light chain (isoform 1) - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-1995  
C:Accession: S47997  
R:Fan, J.; Amos, L.A.  
J. Mol. Biol. 240, 507-512, 1994  
A:Title: Kinesin light chain isoforms in Caenorhabditis elegans.  
A:Reference number: S47997; MUID:94322393  
A:Accession: S47997  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-563 <FAN>  
A:Cross-references: EMBL:Z29644

Query Match 48.0%; Score 1517; DB 2; Length 563;  
Best Local Similarity 56.7%; Pred. No. 9.1e-81;  
Matches 317; Conservative 75; Mismatches 125; Indels 42; Gaps 9;

QY 17 LSQEEILGSTRVLSQGLEALRSEHQAVLQSLQTECLQGGHVEGHEKARQLRRSME 76  
D 4 MSQDDVTGLRTVQOGLAEALREHSTISNTLETSTVKGVE--DEAPLPKQLSQINDNLD 61  
QY 77 NIELGLSEAQVMLALASHLSTVESEKQLRAQVRLCQENQWLRLDELACTQORLORSEQA 136  
D 62 KLVCVGVDETSLMLVFQLTQGMADAHQYQARRRRCQENAWLRDELSTQIKLQSQEQM 121  
QY 137 VAQLEEKHLEFLGQLROYDEDDGHTSEEK-----DATKDSLDDL-FPNEEEED--- 186  
D 122 VAQLEENKHLKYMASIKQLD-DGTQSDTKTSVDVGPQVPTNETLQELGFGPEDEEDMNA 180  
QY 187 -----PSNGLSRGQGATAAQGGYEIPARLRTLNLVLIQYAAQGRYEVAVPLCKQALE 239  
D 181 SQFNQPTPANQMA-----ASANVGYEIPARLRTLNLVLIQYASQGRYEVAVPLCKQALE 234  
QY 240 DLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPAVALTN 299  
D 235 DLEKTSRHDHPDVATMLNIALVYRDQNKYKEAANLLNEALSIREKCLGESHHPAVALTN 294  
QY 300 NLAVLYGKRGKYKEAEPICQRALEIREKVLGTNHPDVAKQLNNLALCQNQGYEAEVY 359  
D 295 NLAVLYGKRGKYKEAEPICQRALEIREKVLGTNHPDVAKQLNNLALCQNQGYEAEVY 354  
QY 360 YQALAIYEGQLGPDNPVARTKNNLASCYLKQGYAEATLYKEILTRAHVQFEGSVDD 419  
D 355 YKRALEIYESKLGPDNPVARTKNNLSSAYLKQGYKEAEELYKQILTRAHEREFGQISG 414  
QY 420 DHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWKYKACKVSSPTVNTTLNLGALYRRQK 479



```

Db      415  ENKPIWQIAEERE---NKHKGEATANEQAGWAKAAKVDSPTVTTILKNLGALYRRQK  471
QY      480  LEAAETLEECALRRQ-----GTDPISQTKVAELLGESDG---RRTSQEGPGDS  526
Db      472  YEAATLEDVALRAKKQHEPLRSGAMGGIDEMSQSMMASTIGGSRNSMTTSTSQMSQMM  531
QY      527  VKFEGGEDASVAVEWSGDG  545
Db      532  ASTIGGSRNSMTTSTSQTG  550

```

```

RESULT 11
S41864
kinesin light chain (isoform 1) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
C:Accession: S41864
R:Fan, J.; Amos, L.A.
submitted to the EMBL Data Library, January 1994
A:Description: Kinesin light chain isoforms in C. elegans.
A:Reference number: S41864
A:Accession: S41864
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-540 <FAN>
A:Cross-references: EMBL:229644; NID:g452601; PID:g452602

```

Query Match	48.0%;	Score 1516;	DB 2;	Length 540;
Best Local Similarity	58.4%;	Pred. No. 9.9e-81;		
Matches 313:	Conservative	74;	Mismatches 107;	Indels 42;
				Gaps 9;

QY	77	NIELGLSEAQVMLALASHLSTVESEKOKLRAQVRRRLCQENQWLDELAGTQORLORSEQA	136
Db	62	KLVCGVDETSMLMVFLQTQGDMAQHOKYQAQRRRLCQENAWLRDELSSQIKLQSQEQM	121
QY	137	VAQLEEEKKHLEFLCQLROYDEDGHTSEEKG-----DATKDSLDDL-FPNEEED---	186
Db	122	VAQLEENKHLKYMASIKQLD-DGTQSDTKTSVDVGPQPVYNETLQELGFGPDEEDMNA	180
QY	187	-----PSNGLSRGGATAAQGGYEIPARLRTLHNLVIQYAAQGRYEVAVPLCKQALE	239
Db	181	SQFNQPTPANQMA-----ASANVGYEIPARLRTLHNLVIQYASQGRYEVAVPLCKQALE	234
QY	240	DLERTSGRHGPDVATMLNILALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHAPAAATLN	299
Db	235	DLEKTSGHDPDVATMLNILALVYRDQNKYKEAANLLNEALSIREKCLGESHPAPAAATLN	294
QY	300	NLAVLYGKRGRKYKEAEPJCQRALEIREKVLGTHNHPDVAKQLNNLALLCQNOGKYEAVERY	359
Db	295	NLAVLFGKRGRKFKAEPPLCKRALEIREKVLGDDHPDVAKQLNNLALLCQNOGKYEEVEKY	354
QY	360	YQRALAIYEGQLGPDNPVARTKNNLASCYLKQGYAEAEETLYKEILTRAHVQEGSVDD	419
Db	355	YKRALEIYESKLGPDPNVAKTNNLSSAYLKQGYKEAEELYKQILTRAHEREFQISG	414
QY	420	DHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWYKACKVSSPTVNTTLRNLGALYRQGK	479
Db	415	ENKPIWQIAEEREE---NKHKGEGATANEQAGWAKAAKVDSPVTTLKNLALYRQGK	471
QY	480	LEAAETILEECALRSRRQ-----GTDPISQTKVAELLGSDG---RRTSQEG	522
Db	472	YFAAETTFDVALPAKAKOHEPTIRSGAMGGIDEMSOSSMASTIGGSRNSMTTSTQSG	527

RESULT 12  
S41865  
kinesin light chain (isoform 2) - *Caenorhabditis elegans*

C;Species: Caenorhabditis elegans  
C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997  
C;Accession: S41865  
R;Fan, J.; Amos, L.A.  
submitted to the EMBL Data Library, January 1994  
A;Description: Kinesin light chain isoforms in C. elegans.  
A;Reference number: S41864  
A;Accession: S41865  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-522 <FAN>  
A;Cross-references: EMBL:229645; NID:q452603; PID:q452604

Query Match	47.2%;	Score 1491.5;	DB 2;	Length 522;
Best Local Similarity	60.8%;	Pred. No. 2.5e-79;		
Matches 302; Conservative	69;	Mismatches 97;	Indels 29;	Gaps 7;

QY	77	NIELGLSEAQMVLALASHLSTVESEKQKLRAQVRRILCQENQWLRLDELACTQORLQRSEQA	136
Dd	83	KLVCGVDETSLMLVMFQLTQGMDAQHQKYQAQRRRLCQENAWLRDELSSTQIKLQQSEQM	142
QY	137	VAQLEEEKHLEFLGQLROYDEDGHTSEEKEG-----DATKDSLDDL-FPNEEEED---	186
Dd	143	VAQLEEFENKHLKYMASIKQLD-DGTQSDTKTSVDVGPPVTNETLQELGFPEDEEDMNA	201
QY	187	-----PSNGLSRGGGATAACQGYEIPARLRTLNLVLVIQYAAQGRYEAVPLCKQALE	239
Dd	202	SQFNQPTPANOMA-----ASANVGYEIPARLRTLNLVLVIQYASQGRYEAVPLCKQALE	255

QY	240	DLETSRGHPDVATMLNILALVYRDQNKYKEAAHLLNDALSIRESTLGPDPAPVAATLN	299
		:	
Db	256	DLEKTSGHDPDVATMLNILALVYRDQNKYKEAANLNEALSIREKCLGESHAPVAATLN	315
		:	
QY	300	NLAVLYGRGKYKEAEPLCQRALEREKVLTGNHPDPAQLNNLALLCQNQGKEYEAVERY	359
		:             :	
Db	316	NLAVLFGRKGFKDAEPLCKRALEREKVLTGDDHPDPAQLNNLALLCQNQGKEYEEVEKY	375
		:             :	
QY	360	YORALAIYEGQLGPDNPVARTKNNLASCYLKQGVAFEAETLYKEILTRAHVQEFGSVDD	419
		:          :                                     :       :	
Db	376	YKRALEIYESKLGPDPNVANTKNLSSAYLKQGVKEAEELYKQILTTRAHERFEGQISG	435
QY	420	DHKPIWMHAEREEMSKSRHHGGTPYAEYGGWYKACKVSSPTVTTLRLNGLALYRRQOK	479
		:                  :	
Db	436	ENKPWIQAEREE--NKHKGEATANEQAGWAKAAKVDSPTVTTLTKNLGALYRRQOK	492
QY	480	LEAAETLEECALRSRRO	496
		:       :       :	
Db	493	YEAAETLEDVALRAKKQ	509

RESULT 13

S33816  
kinesin light chain isoform 4 - sea urchin (Strongylocentrotus purpuratus) .  
C/Species: Strongylocentrotus purpuratus (purple urchin)  
C/Date: 08-Dec-1993 #sequence\_revision 01-Dec-1995 #text\_change 09-Sep-1997  
C/Accession: S33816  
R/Wedaman, K.P.; Knight, A.E.; Kendrick-Jones, J.; Scholey, J.M.  
J. Mol. Biol. 231, 155-158, 1993  
A/Title: Sequences of sea urchin kinesin light chain isoforms.  
A/Reference number: S33813; MUID:93267648  
A/Accession: S33816  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-451 <WED>

Query Match 46.08; Score 1453; DB 2; Length 451;



Best Local Similarity 66.5%; Pred. No. 3.5e-77;  
Matches 300; Conservative 51; Mismatches 80; Indels 20; Gaps 7;

QY	1	MSGLVLGQDEPAGHR--LSQEEILGSTRIVSQGLEALRSEHQAVLQSLSQTIETCLQQG-	57
		:      :      :     :      :     :      :     :      :     :	
Db	1	MSGSKLSTPNNSGGGQNLNQEQIITGTREVIKLEQLKNEHNDILNSLYQSLKMLKKDT	60
QY	58	GHEEGLVHEKARQLRRSMENIELGLSEAQVMLALASHLSTVSEKQKLRQAVRRRLCQENQ	117
		:       : :  :                   :	
Db	61	PGDSNLVEKTDIIKESLESLELGLGEAKVMALGHHLNMVEAEKQKLRQAVRRRLVQENT	120
QY	118	WLRDELAGTQORLQRSEQAVALKEEKHHLEFLGQLRQYDEDEGHTSEEKGDAKDSLDD	177
Db	121	WLRDELAATQKQLTSEQNLADLEVYKHKHLEYMNSIKKYDEDRTPDEEA---SSSDPLDL	177
QY	178	LFPNEEEEDPSNGLS-----RQGATAAQGGYEIPARLRTLHNLVVIQYAAQGRYEV	229
		:    :	
Db	178	GFP----EDDDGGQADESYPOPQTGSGSVSAAAGGYEIPARLRTLHNLVVIQYASQRYEV	233
QY	230	AVPLCKQALEDLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLNDALSIRESTLGP	289
Db	234	AVPLCKQALEDLERTSGHDPDVATMLNIALVYRDQNKYKEAGNLLHDALAIREKTLGP	293
QY	290	DHPAVAATLNNLAVLYGKRGYKEAEPLCQRALEIREKVLGTNHPDPAKQLNNLALLCQN	349
Db	294	DHPAVAATLNNLAVLYGKRGYKEAEPLCKRALEIREKVLGKDHPDPAKQLNNLALLCQN	353
QY	350	QGYEAVERYQRALAIYEGOLGPDNPVARTKNNLASCYLKQGYAEATLYKEILTRA	409
Db	354	QGYEEVEWYQRALEIYKGLGPDNPVAKTKNNLAAAYLKQGYKAAETLYKQVLTTRA	413
QY	410	HVQEFQ-SVDD-DHKPIWMHAEEREEMSKSR	438
		:	
Db	414	HEREFGLSADDKDNKPIWMQAEEREKVKRR	444
RESULT 14			
T23827			
hypothetical protein M7.2 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999			
C:Accession: T23827			
R:Lennard, N.			
submitted to the EMBL Data Library, January 1996			
A:Reference number: Z19804			
A:Accession: T23827			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-516 <WIL>			
A:Cross-references: EMBL:Z68337; PIDN:CAA92746.2; GSPDB:GN00022; CESP:M7.2			
A:Experimental source: clone M7			
C:Genetics:			
A:Gene: CESP:M7.2			
A:Map position: 4			
A:Introns: 40/1; 145/3; 176/1; 440/2			

	Query Match	30.5%	Score	965;	DB 2;	Length	516;			
	Best Local Similarity	43.8%	Pred.	No. 8.3e-49;						
	Matches	217;	Conservative	80;	Mismatches	172;	Indels	26;	Gaps	9;
QY	20	EEILGSTRLVSQGLEALRSEHQAVQLSLSQTIECQQGGHHEGLVHEKARQLRRSMENIE	79							
Db	7	DDITTVLKTQQTLFALRDDEHAATRILEANL--INSDSSEPSLPSEKMGGLIDESLGKVM	64							
QY	80	LGLSEAQVMLALASHLSTVESEKQKLRQAQVRRRLCQENQWLRLDELAGTQQORLQRSEQAQVAQ	139							
Db	65	DGGDEASLLIMMDKLMQSDVQLSKNHESIRLLRQENTWLWLDLTTTQKRLQESERTVAH	124							
QY	140	LEEEKKHLEFLGQLRQYDED-GHTSEEKEGDATKDSLDDLPNNEEEDPNSGLSRQGQAT	198							
Db	125	LEERDHYKFQDSMNYLNSDFQHTTTSVDATPMWDTLQELGFGPEEDQNNN-QADQGCQR	183							

QY	199	AAQ-----QGGEIPARLRTLHLNLVIQYAAQGRYEVAVPLCKQALEDLERTSGRHPDVA	255
		::::                                  :	
Db	184	SSFSNPISNDYQLPTRLQTQLNVLVIQYMEQGRFEVAIPLCKQALEDVVKVHGVNHL DVA	243
		::::                                  :	
QY	254	TMLNILALVYRDQNKYKEAAHLLNDALSIRESTLGPDHPAVAATLNNLAVLYCKRGKYKE	313
		::::                                  :	
Db	244	TMLNVLAIVRNQENFKDAIYLEKALSIRVQCCGENHHLSVAATLNNLAIAYGKRKYKE	303
		::::                                  :	
QY	314	AEP LQRAL EIREKVLGTNHPDVAKQLNNLALLCQNQGYEAVERYQORALAIYEGQLGP	373
		::::                                  :	
Db	304	SEPLCKRALEIRKNLLGPNHPDVAKQLTNLGIVTQOLEKYETEENYFKQALSINRAFPE	363
		::::                                  :	
QY	374	DNP NVARTKNNLASCYLKQGYAEATLYKEILTRAHVQEFSGVD DDHKPIWMHAERE EE	433
		::::                                  :	
Db	364	NHQNVIKTRKNQLASVFLKQGYQEAEEMYNILSKVAI-----TG NKPIWRIAEDREE	416
		::::                                  :	
QY	434	MSKSRHHEGGTPYAIEYGGW-YKACKVSSPTVTTLRLNLGALYRRQGLKEAAETLEEALR	492
		::::                                  :	
Db	417	RQRN----GIPKVDDSEFVNPTIVMDSNMSTIKNLAAVYRKQKGKEEAAGTLEE-ALG	470
		::::                                  :	
QY	493	SRRQ---GTDPI SQT	504
		::::                                  :	
Db	471	AKKQINGGADHTNST	485
		::::                                  :	
 RESULT 15 S47998			
kinesin light chain (isoform 2) - Caenorhabditis elegans			
C;Species: Caenorhabditis elegans			
C;Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-1999			
C;Accession: S47998			
R;Fan, J.; Amos, L.A.			
J. Mol. Biol. 240, 507-512, 1994			
A;Title: Kinesin light chain isoforms in Caenorhabditis elegans.			
A;Reference number: S47997; MUID:94322393			
A;Accession: S47998			
A>Status: preliminary			
A:Molecule type: mRNA			
A;Residues: 1-114 <FAN>			
A;Cross-references: EMBL:229645			

Query Match 7.6%; Score 241.5; DB 2; Length 114;  
Best Local Similarity 47.8%; Pred. NO. 9.8e-08;  
Matches 54; Conservative 14; Mismatches 32; Indels 13; Gaps 2;

QY	384	NLASCYLKOGKYAAEAETLYKEILTRAHVQEFSGVDDHKKPIWMHAEEEREEMSKSRHHEGG	443
D6	2	NLAGSTIQAYRQRKIESLAK-----MSNMQDDVPWQIAEEREE---NKHKGE	48
QY	444	TPYAEGGWYKACKVSSPTVTTLRLNGALYRRQGKLEAETLEECALRSRQ	496
D6	49	ATANEQAGWAKAAKVDSPTVTTLKNLGALYRRQGYEAETLEDVALRAKKQ	101

Search completed: August 15, 2000, 10:55:49  
Job time: 23076 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2000, 10:52:23 ; Search time 42.09 Seconds  
(without alignments)  
455.744 Million cell updates/sec

Title: US-09-036-614A-1

Perfect score: 3161

Sequence: 1 MSGLVLGQRDEPAGHRLSQE.....APLOVSRGLSASMTDLSSSS 619

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2000	63.3	569	1 KNLC_HUMAN	Q07866 homo sapien
2	1985.5	62.8	556	1 KNLC_RAT	P37285 rattus norv
3	1821	57.6	686	1 KNLC_STRPU	Q05090 strongyloce
4	1759.5	55.7	571	1 KNLC_LOLPE	P46825 loligo peal
5	1709	54.1	508	1 KNLC_DROME	P46824 drosophila
6	1516	48.0	540	1 KNLC_CAEEL	P46822 caenorhabdi
7	179.5	5.7	670	1 SR72_HUMAN	O76094 homo sapien
8	179	5.7	1966	1 MYSB_CAEEL	P02566 caenorhabdi
9	178.5	5.6	1321	1 IF3X_DICDI	O15818 dictyosteli
10	175.5	5.6	670	1 SR72_CANFA	P33731 canis famli
11	173.5	5.5	848	1 MYSP_DIRIM	P13392 dirofilaria
12	171.5	5.4	880	1 MYSP_BRUMA	Q01202 brugia mala
13	171.5	5.4	1857	1 MYST_HUMAN	P35749 homo sapien
14	169.5	5.4	879	1 MYSP_ONCVO	Q02171 onchocerca
15	166	5.3	1972	1 MYST_RABIT	P35748 oryctolagus
16	162.5	5.1	1300	1 DYN_A_NEUCR	Q01397 neurospora
17	162	5.1	1969	1 MYSA_CAEEL	P12844 caenorhabdi
18	160	5.1	1134	1 IF3X_HUMAN	O75153 homo sapien
19	160	5.1	1920	1 PCNT_MOUSE	P48725 mus musculu
20	159.5	5.0	543	1 IEFS_HUMAN	P31948 homo sapien
21	158.5	5.0	3321	1 KEND_HUMAN	O95613 homo sapien
22	157.5	5.0	866	1 MYSP_SCHMA	P06198 schistosoma
23	155.5	4.9	1036	1 OGT1_RAT	P56558 rattus norv
24	155.5	4.9	1505	1 CDP_HUMAN	P39880 homo sapien
25	154	4.9	619	1 OM70_NEUCR	P23231 neurospora
26	153.5	4.9	882	1 MYSP_CAEEL	P10567 caenorhabdi
27	153	4.8	1938	1 MYSD_CAEEL	P02567 caenorhabdi
28	152.5	4.8	1325	1 G160_MOUSE	P35937 mus musculu
29	152	4.8	820	1 FES_FELCA	P14238 felis silve
30	151.5	4.8	920	1 OGT1_HUMAN	O15294 homo sapien
31	151.5	4.8	1960	1 MYSN_HUMAN	P35579 homo sapien
32	151	4.8	576	1 PEX5_PICPA	P33292 pichia past
33	150.5	4.8	1597	1 CTRO_MOUSE	P49025 mus musculu

34	149.5	4.7	955	1 KINL_LEICH	P46865 leishmania
35	149.5	4.7	1805	1 NEST_RAT	P21263 rattus norv
36	149	4.7	1947	1 MYSC_CAEEL	P12845 caenorhabdi
37	148.5	4.7	1938	1 MYS_AEQIR	P24733 aequipecten
38	148	4.7	1509	1 MYSN_ACACA	P05659 acanthamoeb
39	147	4.7	3210	1 CENF_HUMAN	P49454 homo sapien
40	146.5	4.6	866	1 MYSP_SCHJA	Q05870 schistosoma
41	146.5	4.6	2411	1 MYSA_DROME	P05661 drosophila
42	146	4.6	1978	1 MYSG_CHICK	P10587 gallus gall
43	145.5	4.6	1940	1 MYSE_HUMAN	P11055 homo sapien
44	144.5	4.6	1155	1 IF3X_SCHPO	O59742 schizosacch
45	144	4.6	1928	1 MYS1_YEAST	P08964 saccharomyc

ALIGNMENTS

RESULT 1  
KNLC\_HUMAN  
ID KNLC\_HUMAN STANDARD; PRT; 569 AA.  
AC Q07866;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE KINESIN LIGHT CHAIN (KLC).  
GN KLC1 OR KLC OR KNS2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94099888.  
RA Cabeza-Arvelaiz Y., Shih L.-C.N., Hardman N., Asselbergs F.,  
RA Bilbe G., Schmitz A., White B., Siciliano M.J., Lachman L.B.;  
RT "Cloning and genetic characterization of the human kinesin  
RT light-chain (KLC) gene."  
RL DNA Cell Biol. 12:881-892(1993).  
CC -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING  
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. THE LIGHT  
CC CHAIN MAY FUNCTION IN COUPLING OF CARGO TO THE HEAVY CHAIN OR IN  
CC THE MODULATION OF ITS ATPASE ACTIVITY.  
CC -!- SUBUNIT: OLIGOMERIC COMPLEX COMPOSED OF TWO HEAVY CHAINS AND  
CC TWO LIGHT CHAINS.  
CC -!- ALTERNATIVE PRODUCTS: AS FOR THE RAT KINESIN LIGHT CHAIN, THE  
CC HUMAN PRODUCT MAY WELL UNDERGO ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: FOUND IN A VARIETY OF TISSUES. MOSTLY ABUNDANT  
CC IN BRAIN AND SPINE.  
CC -!- DOMAIN: THE LIGHT CHAIN IS COMPOSED OF THREE STRUCTURAL DOMAINS: A  
CC LARGE GLOBULAR N-TERMINAL DOMAIN WHICH MAY BE INVOLVED IN BINDING  
CC TO KINESIN HEAVY CHAINS, A CENTRAL ALPHA-HELICAL COILED-COIL  
CC DOMAIN THAT MEDIATES THE LIGHT CHAIN DIMERIZATION; AND A SMALL  
CC GLOBULAR C-TERMINAL WHICH MAY PLAY A ROLE IN REGULATING  
CC MECHANOCHEMICAL ACTIVITY OR ATTACHMENT OF KINESIN TO MEMBRANE-  
CC BOUND ORGANELLES.  
CC -!- SIMILARITY: BELONGS TO THE KINESIN LIGHT CHAIN FAMILY.  
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
EMBL; L04733; AAA16576.1; -.  
MIM; 600025; -.  
PFAM; PF00409; kinesin2; 5.  
PRINTS; PRO0381; KINESINLIGHT.  
PROSITE; PS01160; KINESIN\_LIGHT; 4.  
Motor protein; Microtubules; Coiled coil; Repeat;  
Alternative splicing.  
DOMAIN 45 150  
DOMAIN 234 401  
COILED COIL.  
4 X 42 AA APPROXIMATE TANDEM REPEATS.

```
FT REPEAT 234 275 1.
FT REPEAT 276 317 2.
FT REPEAT 318 359 3.
FT REPEAT 360 401 4.
SQ SEQUENCE 569 AA; 64786 MW; 5D9376C3EEDD00FE CRC64;

Query Match 63.3%; Score 2000; DB 1; Length 569;
Best Local Similarity 69.8%; Pred. No. 3.5e-105;
Matches 395; Conservative 77; Mismatches 76; Indels 18; Gaps 8;

QY 1 MSGLVLQORDEPAGHRLSQEEILGSTRVLSQGLEALRSEHQAVLSQIECLQQGGHE 60
Db 1 MSTVMYIKEDKL--EKLTDQDEIIISKTQVIOGLEALKNEHNSILQSLLETLCKLKK-DDE 57

QY 61 EGLVHEKARQLRRSMENIELGLSEAQVMLALASHLSTVESEKQKLRQVRRLCQENQWLR 120
Db 58 SNLVEEKSNMIRKSLEMLELGLSEAQVMMALSNHNLNAVESEKQKLRQVRRLCQENQWLR 117

QY 121 DELAGTQORLQORSEQVAQLEEEKKHLEFLGQLRQYDEDGHTSEEKGDATKDSLDDLFP 180
Db 118 DELANTQKQKSEQVAQLEEEKKHLEFMNQLKKYDDDISPSEDKDSTKEPLDDLFP 177

QY 181 NEEEDPSNGLSR--GOGATAAQGGYEIPARLRTLNLNVIQYAAQGRYEVAVPLCKQAL 238
Db 178 N-DEDDPGQGIQQHSSAAAAAQGGYEIPARLRTLNLNVIQYASQGRYEVAVPLCKQAL 236

QY 239 EDLERTSGRHPDPAVATMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPAVAL 298
Db 237 EDLEKTSKSHDHPDPAVATMLNIALVYRDQNKYKDAANLLNDALAIREKTLGKDHAPAVAL 296

QY 299 NNLAVALYKGRGKYKEAEPLCQRALEIREKVLGTNHPDPAVAKQLNNLALCQNGKYEAVER 358
Db 297 NNLAVALYKGRGKYKEAEPLCQRALEIREKVLGKDHDPDAVAKQLNNLALCQNGKYEEVEY 356

QY 359 YYQRAALAIYEGQLGPDNPNVARTKNNLASCYLKQKGYAEATLYKEILTRAHVQFSGVD 418
Db 357 YYQRALEIYQTKLGPDPNVAKTKNNLASCYLKQKFKQAEATLYKEILTRAHEREFGSVD 416

QY 419 DDHKPIWMHAEEREEMSKSRHHEGTPYAEYGGWKACKVSSPTVNTLRNLGALYRRQG 478
Db 417 DENKPIWMHAEERECKGKQ--KDGTSFGEYGGWKACKVDSPTVTTLKNLGALYRRQG 474

QY 479 KLEAAETLEECALRSRQGTDPISQTKVAELLGESDG---RRTSQEGPGDSVKFE---G 531
Db 475 KFEAAETLEEAAMRSRKQGLDNVHKQORVAEVLNDPENMEKRRSRRESLNVVVKYESGPDG 534

QY 532 GEDASVAVESWGDGSGTLQORSGSLGK 557
Db 535 GEEVSMVSVEWNGVSG---RASFCGK 557

RESULT 2
KNLC_RAT
ID KNLC_RAT STANDARD; PRT; 556 AA.
AC P37285;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE KINESIN LIGHT CHAIN (KLC).
GN KLC1 OR KLC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 92052221.
RA Cyr J.L., Pfister K.K., Bloom G.S., Slaughter C.A., Brady S.T.;
RT "Molecular genetics of kinesin light chains: generation of isoforms
RL by alternative splicing.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10114-10118(1991).
CC -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
```

```
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. THE LIGHT
CC CHAIN MAY FUNCTION IN COUPLING OF CARGO TO THE HEAVY CHAIN OR IN
CC THE MODULATION OF ITS ATPASE ACTIVITY.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX COMPOSED OF TWO HEAVY CHAINS AND
CC TWO LIGHT CHAINS.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST THREE FORMS (A,B AND C) ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THEY ONLY DIFFER IN THE C-
CC TERMINUS. THE SEQUENCE OF FORM C IS SHOWN HERE.
CC -!- DOMAIN: THE LIGHT CHAIN IS COMPOSED OF THREE STRUCTURAL DOMAINS: A
CC LARGE GLOBULAR N-TERMINAL DOMAIN WHICH MAY BE INVOLVED IN BINDING
CC TO KINESIN HEAVY CHAINS, A CENTRAL ALPHA-HELICAL COILED-COIL
CC DOMAIN THAT MEDIATES THE LIGHT CHAIN DIMERIZATION; AND A SMALL
CC GLOBULAR C-TERMINAL WHICH MAY PLAY A ROLE IN REGULATING
CC MECHANOCHEMICAL ACTIVITY OR ATTACHMENT OF KINESIN TO MEMBRANE-
CC BOUND ORGANELLES.
CC -!- SIMILARITY: BELONGS TO THE KINESIN LIGHT CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M75148; -; NOT_ANNOTATED_CDS.
CC EMBL; M75147; -; NOT_ANNOTATED_CDS.
CC EMBL; M75146; -; NOT_ANNOTATED_CDS.
CC PFAM; PF00409; kinesin2; 5.
CC PRINTS; PR00381; KINESINLIGHT.
CC PROSITE; PS01160; KINESIN_LIGHT; 4.
KW Motor protein; Microtubules; Coiled coil; Repeat;
KW Alternative splicing.
FT DOMAIN 45 150 COILED COIL.
FT DOMAIN 234 401 4 X 42 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 234 275 1.
FT REPEAT 276 317 2.
FT REPEAT 318 359 3.
FT REPEAT 360 401 4.
FT VARSPLIC 538 546 MISSING (IN ISOFORM B).
FT VARSPLIC 538 556 VSMVSVEWNGMRKMKGLVK -> A (IN ISOFORM A).
SQ SEQUENCE 556 AA; 63247 MW; 1A843B34BCDDDB53E CRC64;

Query Match 62.8%; Score 1985.5; DB 1; Length 556;
Best Local Similarity 70.2%; Pred. No. 2.2e-104;
Matches 388; Conservative 76; Mismatches 70; Indels 19; Gaps 7;

QY 16 RLSQEEILGSTRVLSQGLEALRSEHQAVLSQIECLQQGGHEGLVHEKARQLRRSM 75
Db 14 KLTQDEIIISKTQVIOGLEALKNEHNSILQSLLETLCKLKK-DDESNLVEEKSSMIRKSL 72

QY 76 ENIELGLSEAQVMLALASHLSTVESEKQKLRQVRRLCQENQWLRDELQAGTQORLQORSEQ 135
Db 73 EMLELGLSEAQVMMALSNHNLNAVESEKQKLRQVRRLCQENQWLRDELQANTQOKLQKSEQ 132

QY 136 AVAQLEEEKKHLEFLGQLRQYDEDGHTSEEKGDATKDSLDDLFPNEEEDPSNGLSR-- 193
Db 133 SVAQLEEEKKHLEFMNQLKKYDDDISPSEDKDSDSSKEPLDDLFPN-DEDDPGQGIQQH 191

QY 194 GOGATAAQGGYEIPARLRTLNLNVIQYAAQGRYEVAVPLCKQALELERTSGRHPDVA 253
Db 192 SSAAAAAQGGYEIPARLRTLNLNVIQYASQGRYEVAVPLCKQALELERTSGHDPDVA 251

QY 254 TMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPAVALTNLNLAVLYKGRGKYKE 313
Db 252 TMLNIALVYRDQNKYKDAANLLNDALAIREKTLGRDHPAVALTNLNLAVLYKGRGKYKE 311

QY 314 APPLCQRALEIREKVLGTNHPDPAVAKQLNNLALCQNGKYEAVERYYQRAALAIYEGQLGP 373
Db 312 APPLCKRALEIREKVLGKDHDPDAVAKQLNNLALCQNGKYEEVEYYQRALEIYQTKLGP 371

QY 374 DNPVNARTKNNLASCYLKQKGYAEATLYKEILTRAHVQFSGVDDHKPIWMHAEERE 433
```



Db 372 DDPNVAKTKNNLASCYLKQKFKQAEITLYKEILTRAHEREFGSVDDENKPIWMAEERE 431  
QY 434 MRSRHHGGTPYAEYGGWKACKVSSPTVNTTLNLGALYRRQKLEAAETLEECALRS 493  
Db 432 CRGKQ--KDGSSFGYGGWKACKVDSPTVTTTLNLGALYRRQKFEAAETLEEAALRS 489  
QY 494 RRGOTDPISTKVAELLGSDG--RRTSQEGPGDSVKFE---GGEDASVAVESWGDGS 546  
Db 490 RKQGLDNVHKQVAEVLNDPENVEKRRSRRESLNDVVKYESGPDGGEVSMVWNG--- 546  
QY 547 GTLQSGSLGKIR 559  
Db 547 ---MRKMKLGLVK 556

RESULT 3  
KNLC\_STRPU STANDARD; PRT; 686 AA.  
ID Q05090; Q05089; Q05088; Q04801;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE KINESIN LIGHT CHAIN (KLC).  
OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;  
OC Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;  
OC Strongylocentrotus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EGG;  
RX MEDLINE; 93267648.  
RA Wedaman K.P., Knight A.E., Kendrick-Jones J., Scholey J.M.;  
RT "Sequences of sea urchin kinesin light chain isoforms";  
RL J. Mol. Biol. 231:155-158(1993).  
CC -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING  
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. THE LIGHT  
CC CHAIN MAY FUNCTION IN COUPLING OF CARGO TO THE HEAVY CHAIN OR IN  
CC THE MODULATION OF ITS ATPASE ACTIVITY.  
CC -!- SUBUNIT: OLIGOMERIC COMPLEX COMPOSED OF TWO HEAVY CHAINS AND  
CC TWO LIGHT CHAINS.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR FORMS (KLC-1 TO KLC-4) ARE  
CC PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE OF FORM 3 (KLC-3)  
CC IS SHOWN HERE.  
CC -!- DOMAIN: THE LIGHT CHAIN IS COMPOSED OF THREE STRUCTURAL DOMAINS: A  
CC LARGE GLOBULAR N-TERMINAL DOMAIN WHICH MAY BE INVOLVED IN BINDING  
CC TO KINESIN HEAVY CHAINS, A CENTRAL ALPHA-HELICAL COILED-COIL  
CC DOMAIN THAT MEDIATES THE LIGHT CHAIN DIMERIZATION; AND A SMALL  
CC GLOBULAR C-TERMINAL WHICH MAY PLAY A ROLE IN REGULATING  
CC MECHANOCHEMICAL ACTIVITY OR ATTACHMENT OF KINESIN TO MEMBRANE-  
CC BOUND ORGANELLES.  
CC -!- PTM: PHOSPHORYLATION MAY MODULATE THE PROCESS OF MECHANOCHEMICAL  
CC COUPLING.  
CC -!- SIMILARITY: BELONGS TO THE KINESIN LIGHT CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L10235; AAA03059.1; -.  
DR EMBL; L10234; AAA03058.1; -.  
DR EMBL; L10233; AAA03057.1; -.  
DR EMBL; L08258; AAA03060.1; -.  
DR PIR; S33814; S33814.  
DR PIR; S33813; S33813.  
DR PIR; S33816; S33816.  
DR PFAM; PF00409; kinesin2; 5.  
DR PRINTS; PR00381; KINESINLIGHT.  
DR PROSITE; PS01160; KINESIN\_LIGHT; 4.

Motor protein; Microtubules; Coiled coil; Repeat;  
KW Alternative splicing; Phosphorylation.  
FT DOMAIN 20 160 COILED COIL.  
FT DOMAIN 240 407 4 X 42 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 240 281 1.  
FT REPEAT 282 323 2.  
FT REPEAT 324 365 3.  
FT REPEAT 366 407 4.  
FT VARSPLIC 564 572 MISSING (IN ISOFORM KLC-2).  
FT VARSPLIC 564 600 MISSING (IN ISOFORM KLC-1).  
FT VARSPLIC 441 451 GKFDNAPYGD -> VKRKPAPAKS (IN ISOFORM  
FT KLC-4).  
FT VARSPLIC 452 686 MISSING (IN ISOFORM KLC-4).  
SQ SEQUENCE 686 AA; 76517 MW; 603D71186CC0364B CRC64;  
  
Query Match 57.6%; Score 1821; DB 1; Length 686;  
Best Local Similarity 57.2%; Pred. No. 4 6e-95;  
Matches 395; Conservative 82; Mismatches 129; Indels 84; Gaps 16;  
  
QY 1 MSGLVLQGRDEPAGHR--LSQEEILGSTRVLSQGLSEALRSEHQAVLQSLSTIECLQQG- 57  
Db 1 MSGSKLSTPNNSGGQGNLSQEIIITGTREVIKGLQKNEHDILNLSLYQSLKMLKKDT 60  
  
QY 58 GHEGLVHEKARQLRRSMENIELGSEAVMLALASHLSTVESEKQKLAQVRLCQENQ 117  
Db 61 PGDSNLVEEKTDIIEKSLSELELGLGEAKVMALGHLNMVAEKQKLAQVRLVQENT 120  
  
QY 118 WLRDELATQORLQORSEQAVQAQLEEEKKHLEFLGQLRQYDEGDGHTSEEKGDATKSLDD 177  
Db 121 WLRDELAATQOKLQTSQONLADLEVYKHKLEYMNSIKKYDEDRTPDEEA---SSSDPLDL 177  
  
QY 178 LFPNEEEEDPSNGLS-----RQGGATAAQGGVEIPARLRTLNLVIOYAAQGRYEV 229  
Db 178 GFP----EDDDGGQADESYPPQPTQSGSVSAAAGGYEIPARLRTLNLVIOYASQSRYEV 233  
  
QY 230 AVPLCKQALEDLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGP 289  
Db 234 AVPLCKQALEDLERTSGRHPDVATMLNIALVYRDQNKYKEAGNLLHDLAIREKTLP 293  
  
QY 290 DHPAATLNLAVLYGRKGYKEAEPLCQRALEIREKVLGTNHPDVAKQLNLLALCQN 349  
Db 294 DHPAATLNLAVLYGRKGYKEAEPLCQRALEIREKVLGTNHPDVAKQLNLLALCQN 353  
  
QY 350 QGYEAEVYQYRALAIYEGQLGPDNPNVARTKNNLASCYLKQGYAEAEYKEILTRA 409  
Db 354 QGYEAEVYQYRALAIYEGQLGPDNPNVARTKNNLAAAYLKQGYKAAETLYKQVLTRA 413  
  
QY 410 HVQEEG-SVDD-DHKPIWMAEEREEMSKSRHHEGTPYAEYGGWKACKV-----SSPT 462  
Db 414 HERFGLSADDDKDKPIWMAEEREEMSKSRHHEGTPYAEYGGWKACKV-----SSPT 470  
  
QY 463 VNTTLNLGALYRRQKLEAAETLEECALRSRQGTDPISQTKVAELG----- 511  
Db 471 VNTTLNLGALYRRQKLEAAETLEECALRSRQGTDPISQTKVAELG----- 530  
  
QY 512 ----ESDGRRTS---QEGPDSVKFE---GGEDASVAVESW--- 543  
Db 531 AMAKERHRRSSGTPRHGSTEVSVEYKTDGSEEVSGVAVKAKRKAHDRSRIPAGYVEI 590  
  
QY 544 -----DGSGLTQRSGLKIRDLVLRSSSELLVRKLOGTEPRPSSSNMKRAASLNY 593  
Db 591 PRSPPHVLVENGDKLRRSGSLKLRASVRRSRSTKLLNKLKGRE-SDDDDGGMKRASSMSV 649  
  
QY 594 L-----NQPSAAPLOVS-RGLSASTMDLSS 617  
Db 650 LPSRGNDSTPAPIQLSQGRVGSNDLSS 679  
  
RESULT 4  
KNLC\_LOLPE  
ID KNLC\_LOLPE STANDARD; PRT; 571 AA.  
AC P46825;













```

AC 015818;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PUTATIVE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT (EIF-3).
GN CLUA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyostelid; Dictyostelium.
RN [1]
RX MEDLINE; 97352794.
RA Zhu Q., Hulen D., Liu T., Clarke M.;
RT "The cluA-mutant of Dictyostelium identifies a novel class of
RL proteins required for dispersion of mitochondria.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:7308-7313(1997).
CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA (BY SIMILARITY).
CC -!- FUNCTION: REQUIRED FOR PROPER DISPERSION OF MITOCHONDRIA.
CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EIF3 P135 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U49332; AAC26527.1; -.
CC DICTYDB; DD0????; CLUA.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 178 185 POLY-GLN.
FT DOMAIN 562 565 POLY-GLU.
FT DOMAIN 1250 1263 POLY-THR.
FT SEQUENCE 1321 AA; 148917 MW; 94B2E1F83159437E CRC64;
SQ

```

Db 1171 F Y T G -----LSQTANQIKLFKQHQAALAEQDELARLQKEKADQFKKSQPRVSAMPPSLE 1222

Qy 584 NMKRAASLYNLN-OPSAAPLQVSRGLSAST 612

Db 1225 NGSVSELLNYNGKPKKSQSKSKSTNTTT 1254

1 : |||:| : :|: :|

RESULT 10

SR72\_CANFA

ID SR72\_CANFA STANDARD; PRT; 670 AA.

AC P3731;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE SIGNAL RECOGNITION PARTICLE 72 KDA PROTEIN (SRP72).

GN SRP72.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

[1]

RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RP MEDLINE; 93273803.

RX Luetcke H., Pohn S., Ashford A.J., Remus M., Frank R.,

RA Dobberstein B.;

RA "Assembly of the 68- and 72-kD proteins of signal recognition

RT particle with 7S RNA.";

RT J. Cell Biol. 121:977-985(1993).

RL -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE

CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC

CC RETICULUM MEMBRANE. SRP72 BINDS THE 7S RNA ONLY IN PRESENCE OF

CC SRP68. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT INTERACT DIRECTLY WITH

CC THE DOCKING PROTEIN IN THE ER MEMBRANE AND POSSIBLY PARTICIPATE

CC IN THE ELONGATION ARREST FUNCTION.

CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE

CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,

CC SRP19, SRP14 AND SRP9.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -!- DOMAIN: THE C-TERMINUS IS ESSENTIAL FOR THE INTERACTION WITH THE

CC SRP68/7S RNA COMPLEX.

CC -!- SIMILARITY: BELONGS TO THE SRP72 FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collabora

CC between the Swiss Institute of Bioinformatics and the EMBL outstat

CC the European Bioinformatics Institute. There are no restrictions on

CC use by non-profit institutions as long as its content is in no

CC modified and this statement is not removed. Usage by and for comm

CC entities requires a license agreement (See <http://www.isb-sib.ch/anno>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; X67813; CAA48014.1; -.

DR PIR; S32167; S32167.

DR PIR; A40692; A40692.

KW Signal recognition particle; Ribonucleoprotein.

FT INIT\_MET 0 0 PROBABLE.

FT MOD\_RES 1 1 BLOCKED.

FT DOMAIN 551 560 POLY-LYS.

FT DOMAIN 661 664 POLY-LYS.

FT SEQUENCE 670 AA; 74362 MW; D394CC56600B5C3D CRC64;

SQ

```
Db 183 CALIGQQLSQAMKILQKAEDLCRRSLSESDGTEEDPOAELAIHQMAYILOQGRTE 242
QY 271 EAAHLNDALSIRESTLGPDPHPAVALTNLNLAVL-----YKRGKYKEAE-----PLC 318
Db 243 EALQLYNQIIKLPDVG-----LLAVIANIITINKDQNVFDSKKVKLTNAEGVEFKLS 298
QY 319 QRALEIRE--KVLGTHNHPDVAKQLNLLALCQNOQKYEAV-----ERYYPALALA 365
Db 299 KKQLQAIENKALLAMVTNQAEQCRKISASLSQSPHELLPVLIAQAALCREKHQHTKAIE 358
QY 366 IYEGQLGPDNP-NVARTKNNLASCYLKQGYAEAEETLYKEILTRAHVQEFSGSVDDHKP- 423
Db 359 LLQ-EFSDQHPENAAEIKLTMALQKISQGNISKACLILRSI-----EEL-----XHKPG 406
QY 424 -----IWMHAEREEMSKSRHHEGGTPYAEYGGWKACKVSSPTVNTTLRNLGALYRROG 478
Db 407 MYSALVTMYSHEEDIDSAIE-----VFTQAIQWYQNHQPKSSAHLSLIREAANFKLYG 460
QY 479 -KLEAAETLEEC-----ALRSRROGTDPISTQKVAEEL 510
Db 461 RKKEAISDLQWLKQNPDKDIHTLAQLISAVSLVDPEKAKALSKHLPSSDSMSLKVDVEAL 520
QY 511 GESDG-----RRTSQEGPGDSVKFEGGE-----DASVAVE---WSGDG 545
Db 521 ENSPGATYIRKKGGKVAGDSQPKQEQGDURKKKKKKKKKGLPKNYDPKVTDPDWLPNR 580
QY 546 SGTIQRSGSLGKIRDVLR-----SSELLVRKLQ-----TEPRPSSNMKRAASLNY 593
Db 581 ERSYRGRKKKKKKKDIQKGTQGTAGASSELDASKTVSSPPTSPPRPGSAATASASTSNI 640
QY 594 L-----NQPSAAP 601
Db 641 IPPRHQKPAGAP 652

RESULT 11
MYSYP_DIRIM
ID MYSYP_DIRIM STANDARD; PRT; 848 AA.
AC P13392;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PARAMYOSIN (FRAGMENT).
OS Dirofilaria immitis.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Dirofilaria.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90220759.
RA Limberger R.J., McReynolds L.A.;
RT "Filarial paramyosin: cDNA sequences from Dirofilaria immitis and
RT Onchocerca volvulus.";
RL Mol. Biochem. Parasitol. 38:271-280(1990).
RN [2]
RP SEQUENCE OF 18-96 FROM N.A.
RX MEDLINE; 89344126.
RA Grandea A.G. III, Tuyen L.K., Asikin N., Davis T.B., Philipp M.,
RA Cohen C., McReynolds L.A.;
RT "A lambda gt11 cDNA recombinant that encodes Dirofilaria immitis
RT paramyosin.";
RL Mol. Biochem. Parasitol. 35:31-41(1989).
CC -!- FUNCTION: PARAMYOSIN IS A MAJOR STRUCTURAL COMPONENT OF
CC MANY THICK FILAMENTS ISOLATED FROM INVERTEBRATE MUSCLES.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29733; AAA28299.1; -.
DR EMBL; J04009; AAA28300.1; -.
DR PIR; A45548; A45548.
DR PIR; A44972; A44972.
DR HSSP; P80220; 1DIP.
DR PFAM; PF01576; Myosin_tail; 1.
KW Coiled coil; Heptad repeat pattern; Muscle protein; Thick filament;
KW Myosin.
FT NON_TER 1 1
FT DOMAIN <1 9 NONHELICAL REGION (POTENTIAL).
FT DOMAIN 10 833 COILED COIL (POTENTIAL).
FT DOMAIN 834 848 NONHELICAL REGION (POTENTIAL).
FT DISULFID 105 105 INTERCHAIN (POTENTIAL).
FT DISULFID 594 594 INTERCHAIN (POTENTIAL).
FT NON_TER 848 848
SQ SEQUENCE 848 AA; 98008 MW; BF32B7120157FA50 CRC64;

Query Match 5.5%; Score 173.5; DB 1; Length 848;
Best Local Similarity 22.0%; Pred. No. 0.0095;
Matches 138; Conservative 109; Mismatches 242; Indels 139; Gaps 29;

QY 34 EALRSEHOAVLQSLSTIECLQGGHHEGLVHEKARQLRRSMENIELGLSEAQVMLALAS 93
Db 125 ERQLQHEVI--ELTATIDQLQDKH---LAERAAE-RFEAQTIEL---SNKVEDLNR 173
QY 94 HLSTVESEKQLRAQVRRRLCOBNQWLRLDELATGQ---QRLQRSEQAVALLEE-EKKHLE 148
Db 174 HVNDLAQQRQLQAENNDLLKEIHDQKVOLDNLQHVKYQLAQOLEEARRLDEAERSQ 233
QY 149 FLGQLROYDEDEGHTSEEKGDATKDSLDDLPNNEEDPNSGLSRGQGAATAAQGGYEIP 208
Db 234 LQAQLHQV-----QLELDSVRTALDEESAARAEAEHKLALAN---TEITQWKSFD 281
QY 209 ARLRTLHNLVIQ-----YAAQGRYEAVPLCKQALDELERTSGRHPDVAATML----- 256
Db 282 AEV-ALHHEEVEDLRKKMLQKQAEYEEQIEIMLQKISQLEKAKSRLOQSEVEVLIVLEKA 340
QY 257 -NILALVYRDQNKYKEAAHLN---DALSIRESTLGPDPHPAVALTNLNLAVLYKRGKYK 312
Db 341 QNTIAILERAKEQLEKTVNELKVRIDELTVELEAAQREARAALAEQKMKNLNLYEKAIEQK 400
QY 313 EA-----EPLCQRALEIREKVLGTHNHPDVAKQLNLLAL-----LCQNOQKYEAVE 357
Db 401 EALARENKKIQLDDLHEAKEALADANRKLHLDLENARLAGEIRELQTLAKSESAARRDAE 460
QY 358 RYQYRALAIYEGQLGPDNPVNARTKNNLASCYLKQGY-----AEAEETLYKEILTR 408
Db 461 NRAQRALAEQ-QLRIMERRLQEEEMEALRKNMQFEIDRLTAALADAEARMKAERIS 519
QY 409 -----AHVQEEFG-SVDDDHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWKACKVSSP 461
Db 520 LKKYQAEIAELEMTVDNLNR---ANIEAQTKIKQS-----EQLKILQA 561
QY 462 TVNTTLRLNLG-----ALYRRQGLKLEAAETLEEC-----ALRSRROG-----T 498
Db 562 SLEDTQRQLQQLDQYALAQKVSALSAB-LEECKVALDNARARKQAEIDLEEANARIT 620
QY 499 DPIS-----QTKVAELLGESDGRRTSQEGPGDSVKFEGGEDASVAVESGDGS 546
Db 621 DLVSINNLTAKNKLETETELSTAQADLD-EATKELHAADERANRALADAARAVEQLHE-- 677
QY 547 GTLQRSGSLGKIRDVLRSSSELLVRKLQ 574
Db 678 ---EQEHSN-KI-DALRKSLEEQVKQLQ 700

RESULT 12
MYSYP_BRUMA
ID MYSYP_BRUMA STANDARD; PRT; 880 AA.
```







```
CC -!- DISEASE: A CHROMOSOMAL REARRANGEMENT, KNOWN AS PERICENTRIC
CC INVERSION INV(16)(P13Q22), PRODUCES A FUSION PROTEIN THAT CONSISTS
CC OF THE 165 N-TERMINAL RESIDUES OF CBF-BETA (PEPB2) WITH THE TAIL
CC REGION OF MYH11. THIS REARRANGEMENT IS ASSOCIATED WITH ACUTE
CC MYELOID LEUKEMIA OF M4EO SUBTYPE.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF001548; AAC31665.1; -.
CC EMBL; D10667; -; NOT_ANNOTATED_CDS.
CC EMBL; X69292; CAA49154.1; -.
CC HSSP; P08287; IGHC.
CC MIM; 160745; -.
CC PFAM; PF00612; IQ; 1.
CC PFAM; PF01576; Myosin_tail; 1.
CC PFAM; PF00063; myosin_head; 1.
CC
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Multigene family;
CC Proto-oncogene; Chromosomal translocation.
CC
CC NON_TER 1
CC DOMAIN <1 728 GLOBULAR HEAD (S1).
CC DOMAIN 729 1819 RODLIKE TAIL (S2 AND LMM DOMAINS).
CC DOMAIN 1820 1857 CARBOXYL-TERMINAL.
CC NP_BIND 63 70 ATP (POTENTIAL).
CC DOMAIN 546 568 ACTIN-BINDING (BY SIMILARITY).
CC DOMAIN 647 661 ACTIN-BINDING (BY SIMILARITY).
CC DOMAIN 670 692 IQ.
CC DOMAIN 729 1819 COILED COIL (POTENTIAL).
CC MOD_RES 14 14 METHYLATION (TRI-) (POTENTIAL).
CC MOD_RES 586 586 ALKYLATION (SH-1) (POTENTIAL).
CC MOD_RES 596 596 ALKYLATION (SH-2) (POTENTIAL).
CC CONFLICT 772 774 EEK -> NSE (IN REF. 2).
CC CONFLICT 1443 1443 T -> S (IN REF. 2).
CC CONFLICT 1495 1496 KQ -> NE (IN REF. 2).
CC CONFLICT 1671 1671 A -> S (IN REF. 3).
CC CONFLICT 1843 1843 T -> L (IN REF. 2).
CC SEQUENCE 1857 AA; 214349 MW; 2F70B17BFE11406 CRC64;

Query Match 5.4%; Score 171.5; DB 1; Length 1857;
Best Local Similarity 20.8%; Pred. No. 0.032;
Matches 162; Conservative 107; Mismatches 289; Indels 219; Gaps 29;

QY 11 EPAGHRLSQEEILGSTRL-----VSQGLEALRSEHQAVLQSL-----Q 49
| | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1000 ELEGHISDLQEDLSERAARNKAQKQKRDLGEELEALKTELEDTLSTATQQLRAKREQ 1059

QY 50 TIECLOQGGHEGLVH-----EKARQLRRSMENIELG---LSEAQV 87
| : : | | | | : | : | : | : | : | : | : | : | :
Db 1060 EVTVLKALDEETRSHEAQVQEMRQKHAQVEELTEQLEQFKRAKALNDKNKQMLEKENA 1119

QY 88 MLA-----LASHLSTVESEKQKLRAQVRRL---CQENQWLRLDELAGTQQLRSEQAV-- 137
| | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1120 DLAGELRVLGQAQVEVHKKKKLEAQVQLQSKSDGERARAEELNDKVHKLQNEVESVTG 1179

QY 138 -----AQL-----EIEKKHLEFLGQLRQYDEDGHTSEEKGD 169
| | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1180 MNEAEGKAIKLAKDVASLSSQLQDTQELLQETRQKLNVTKLRLQLEE-----ERN 1231

QY 170 ATKDSLDDLPNEEEDPSNGLSR-----GQATAAQGGYEIPAR 210
| : | | | | | : | | | | : | | | | : | | | | : | | | | :
Db 1232 SLQDQLD-----EEMAKQNLERHISTLNQLSDSKKKLQDFASTVEALEEGKKRFQ-- 1283
```

```
QY 211 LRTLHLNLIQY---AAQGRYEVAVPLCKQALED-----LERTSGRHPDV 252
: : | | | | : | | | | : | | | | : | | | | :
Db 1284 -KEIENLTQQYEKAAAYDKLTKTNRLQQLDLDLVLDNQRLVSNLEKKQKFDQLL 1342

QY 253 ATMLNILALVYRDQNKYKEA-----AHLNDALSIRESTLGPDPHVAATLNN 300
| | | | | : | | | | : | | | | : | | | | :
Db 1343 AEEKNI-SSKYADERDRAEAAREKETKALSARALEEALEAKEE-LERTNKMKAEMED 1400

QY 301 LAVLYGKRGKYKEAEPLCQAL-----EIREKVLGTNHPDVAKQLNNLALLCQ 348
| | | | | : | | | | : | | | | : | | | | :
Db 1401 LVSSKDDVGKNVHELEKSKRALETQMEEMKTQLEELEDELQATEDAKLRLEVMQALKGQ 1460

QY 349 NQKYEAIVE-----RYQORALAIYEGQLGPDNPVA-----RTKNNLASCYLK--- 391
: : | | | | | : | | | | : | | | | : | | | | :
Db 1461 FERDLQARDEQNEEKRRQLRQLHEYTELEDERKQRALAAAKKKLEGDLKDLQADS 1520

QY 392 --QGYAEAEETLYKEILTRAHVQFEGSVDDDHK---PINMHAEREEMSKSRHHEGTP 445
: | : | : | : | : | : | : | : | : | : | : | :
Db 1521 AIKGREEAIKLRK---LQAQMKDFQLEEDARASRDEIFATAKENEKKAASLEAD---- 1573

QY 446 YAEYGGWYKACKVSSPTVNTTLNLGAL--YRROGKLEAAETLEECAL---RSRRQGTD 499
: : | | | | : | | | | : | | | | : | | | | :
Db 1574 -----LMQLQEDLAAAEARAKQADLEKEELABELASSLSGRNALQDEK 1616

QY 500 PISQTKVAELLGESDGRRTSQEGPGSDSVKFEGGEDASVAVESGDSGTLQRSGLGKIR 559
: : | | | | : | | | | : | | | | : | | | | :
Db 1617 RRLEARIAQLEEELEEEEOGNMEAMSDVRKATQAEQLSNELATERS-TAOKNES---AR 1672

QY 560 DVLRRSSELLVRKLQGTPEPRPSSSNMKRAASLN-YLNQPSAAPLQVSRGLSASTMDL 615
| | : | | | | | | | | | | | | | | | | | | | |
Db 1673 QQLERQNKELRSKLHEMEGAVKSKFKSTIAALEAKIAQLEEQVEQAREKQAATKSL 1729

RESULT 14
MISP_ONCVO
ID MISP_ONCVO STANDARD; PRT; 879 AA.
AC Q02171;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PARAMYOSIN.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93165084.
RA Dahmen A., Gallin M., Schumacher M., Erttmann K.D.;
RT "Molecular cloning and pre-mRNA maturation of Onchocerca volvulus
paramyosin.";
RL Mol. Biochem. Parasitol. 57:335-338(1993).
CC -!- FUNCTION: PARAMYOSIN IS A MAJOR STRUCTURAL COMPONENT OF
MANY THICK FILAMENTS ISOLATED FROM INVERTEBRATE MUSCLES.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M95813; AAA29431.1; -.
CC HSSP; P80220; 1DIP.
CC
CC PFAM; PF01576; Myosin_tail; 1.
CC Coiled coil; Heptad repeat pattern; Muscle protein; Thick filament;
KW Myosin.
FT DOMAIN 1 36 NONHELICAL REGION (POTENTIAL).
FT DOMAIN 37 860 COILED COIL (POTENTIAL).
```

```
FT DOMAIN      861      879      NONHELICAL REGION (POTENTIAL).
SQ SEQUENCE    879 AA; 101125 MW; DCA24AC01A9F02B1 CRC64;

Query Match      5.4%; Score 169.5; DB 1; Length 879;
Best Local Similarity 22.0%; Pred. No. 0.017;
Matches 138; Conservative 107; Mismatches 244; Indels 139; Gaps 29;

QY 34 EALRSEHQAVLSQSLQIECLQQGGHEGLVHEKARQLRRSMENIELGLSEAQVLMALAS 93
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 152 ERQRLQHEVI--ELTAAIDQLQDKH----LAEKAAE-RFEAQTIEL-----SNKVEDLNR 200

QY 94 HLSTVESEKQLRAQVRRRLCQENQWLRLDELGTQ----QRLQRSEQAVQAQLEE-EKKHLE 148
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201 HVNDLAQRQRQAENNDLLKEIHDQKVQDLNLQHVKYQLAQQLSEARRRLEDAERERSQ 260

QY 149 FLGQLROYDEDGHTSEKEGDAKDSLDLFPNEEEDPSNGLSRGQGATAAQGGYEIP 208
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 261 LQAQLHQV-----QLELDSVRTALDEESAARAEAEHKLALAN----TEITQWKSQFD 308

QY 209 ARLRTLNLVIQ-----YAAQGRYEVAVPLCKQALEDLERTSGRHPDVATML----- 256
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 309 AEV-ALHHEEVEDLRKKMLQKQAEYEEQIEIMLQKISQLEKAKSRLQSEVEVLIVDLEKA 367

QY 257 -NILALVYRDQNKYKEAAHLN---DALSIRESTLGPDPAPVAATLNNLNLAVLYGKRGKYK 312
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 368 QNTIALIERAKEQLEKTVNELKVRIDELTVELEAAQREARAALAEQKMKNLVEKAVEQK 427

QY 313 EA-----EPLQORALEIREKVLGTNHPDPAKQLNNLAL-----LCQNOQKYEAVE 357
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 428 EALARENKKLQDDLHEAKEALADANRKLHELLENARLAGEIRELOTALKESAEARRDAE 487

QY 358 RYQQRALAIYEGQLGPDNPVARTKNNLASCYLKQGY-----AAEATLYKEILTR 408
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 488 NRAQRALAELO-QLRIMERRLQKEEEMEALRKNMOPFIDRLTAALADAEARMKAEIAR 546

QY 409 -----AHVQEFQ-SVDDDHKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWKACKVSSP 461
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 547 LKKYQAEIAELEMTVDNLNR---ANIEAQTKIKQS-----EQLKVLQA 588

QY 462 TVNTTLRLNLG-----ALYRROGKLEAAETLEEC-----ALRSRRQG-----T 498
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 589 SLEDTRQLQQTLDQYALAQKVSALSAE-LEECKVALDNAIRARKQAEIDLEEANGRIT 647

QY 499 DPIS-----QTKVAELLGESDGRRTSQEGPGDSVKFEGGEDASVAVEWSGDGS 546
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 648 DLVSINNNLTAIKNKLETELSTAQADLD-EATKELHAADERANRALADARAVEQLHE-- 704

QY 547 GTLQSGSLGKIRDVLRRSSELLVRKLQ 574
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 705 ---EQEHSM-KI-DALRKSLEEQVKQLQ 727

RESULT 15
MYST_RABIT
ID MYST_RABIT STANDARD; PRT; 1972 AA.
AC P35748;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOSIN HEAVY CHAIN, SMOOTH MUSCLE ISOFORM (SMMHC).
GN MYH11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92073350.
RA Babij P., Kelly C., Periasamy M.;
RT "Characterization of a mammalian smooth muscle myosin heavy-chain
RT gene: complete nucleotide and protein coding sequence and analysis of
RT the 5' end of the gene."
RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).
```

Db 1504 RTNKMLKAEMEDLVSSKDDVGKNVHELEKSKRALETOMEEMKTQLELEDELEQATEDAKL 1563

QY 337 AKQLNNLALLCQNGKYEAVE-----RYYQRALAIYEGOLGPDNPNVARTKNNLASCY 389

Db 1564 RLEVNMQALKVQFERDLQARDEQNEEKRRQLQRQLHEYETELEDE-----RKQRALAAAA 1618

QY 390 LK--QGKYAE-----AETLYKEIL-TRAHVQEFGSVDDDHK-----PIWMAEER 431

Db 1619 KKKLEGLDKDLELQADSAIKGREAIKQLLKLOAQMKDFOFELEDARASRDEIFATAKEN 1678

QY 432 EEMSKSRHHEGGTPYAEYGGWYKACKVSSPTVNTTLRNLGAL--YRROGKLEAAFTLEEC 489

Db 1679 EKKAKSLEAD-----LMQLQEDLAAAERARKQADLEKEELAEEL 1717

QY 490 AL----RSRRQGTDPISQTKVAELLGESDGRRTSQEGPGDSVKFEGGEDASVAVEWSGDG 545

Db 1718 ASSLSGRNALQDEKRRLEARIAQLEEELEEEQGNMEAMSDRVRKATQQAEOQLSNELATER 1777

QY 546 SGTLQSGSLGKIRDVLRRSSELLVRKQLQGTPEPRPSSNNMKRAASL 591

Db 1778 S-TAQKNES---ARQQLERQNKELKSKLQEMEGAVKSKFKSTIAAL 1819

Search completed: August 15, 2000, 13:20:24  
Job time: 8881 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2000, 10:46:30 ; Search time 188.37 Seconds  
(without alignments)  
227.838 Million cell updates/sec

Title: US-09-036-614A-1  
Perfect score: 3161  
Sequence: 1 MSGVLVQRDEPAGHRLSQE.....APLQVSRGLSASTMDLSSSS 619

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_12:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_invertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2056.5	65.1	599	11	O88448 mus musculu
2	1910	60.4	541	11	O88447 mus musculu
3	1775	56.2	487	13	Q90707 gallus gall
4	962.5	30.4	521	5	Q21592 caenorhabdi
5	503.5	15.9	490	2	O33965 plectonema
6	276.5	8.7	1049	2	Q9XBP6 myxococcus
7	232	7.3	609	10	O81629 arabidopsis
8	229.5	7.3	886	2	O86768 streptomyc
9	196.5	6.2	707	10	Q9ZW40 arabidopsis
10	189	6.0	342	2	O67735 aquifex aeo
11	189	6.0	948	2	Q55489 synechocyst
12	181	5.7	1939	5	Q25662 plasmodium
13	179.5	5.7	2176	5	O46112 drosophila
14	179	5.7	1963	5	O02244 caenorhabdi
15	178	5.6	1011	5	Q9Y148 drosophila
16	176.5	5.6	1604	10	Q9XI40 arabidopsis
17	167.5	5.3	2101	4	Q14981 homo sapien
18	167.5	5.3	3187	11	Q63714 rattus norv
19	166	5.3	914	10	Q96301 arabidopsis

20	166	5.3	1938	11	O08639	O08639 mus musculu
21	166	5.3	1972	11	O08638	O08638 mus musculu
22	164	5.2	2442	4	O60588	O60588 homo sapien
23	163	5.2	1313	4	O75033	O75033 homo sapien
24	163	5.2	2168	11	Q61043	Q61043 mus musculu
25	162	5.1	1992	5	Q21440	Q21440 caenorhabdi
26	162	5.1	2442	4	O14812	O14812 homo sapien
27	161.5	5.1	1133	5	Q21022	Q21022 caenorhabdi
28	161.5	5.1	2115	4	Q14980	Q14980 homo sapien
29	160.5	5.1	944	10	O82422	O82422 hordeum vul
30	160.5	5.1	1426	4	O60314	O60314 homo sapien
31	160.5	5.1	1756	4	O60454	O60454 homo sapien
32	160	5.1	1134	4	O75153	O75153 homo sapien
33	159.5	5.0	543	11	O54981	O54981 cricetus
34	159	5.0	1974	5	Q21000	Q21000 caenorhabdi
35	158.5	5.0	3321	4	O95613	O95613 homo sapien
36	158	5.0	1743	5	O96063	O96063 dugesia jap
37	158	5.0	1957	5	Q04010	Q04010 onchocerca
38	157.5	5.0	1941	5	Q26079	Q26079 placopecten
39	157.5	5.0	2230	4	Q13439	Q13439 homo sapien
40	157	5.0	886	1	O29230	O29230 archaeoglob
41	157	5.0	1257	4	O75334	O75334 homo sapien
42	157	5.0	1935	5	O44934	O44934 loligo peal
43	156.5	5.0	1390	5	O77033	O77033 dictyosteli
44	156.5	5.0	1958	5	O96062	O96062 dugesia jap
45	156.5	5.0	1999	11	Q63731	Q63731 rattus norv

ALIGNMENTS

RESULT 1  
O88448  
ID O88448 PRELIMINARY; PRT; 599 AA.  
AC O88448;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE KINESIN LIGHT CHAIN 2.  
GN KLC2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB C; TISSUE=BRAIN;  
RX MEDLINE; 98288268.  
RA RAHMAN A., FRIEDMAN D.S., GOLDSTEIN L.S.;  
RT "Two kinesin light chain genes in mice. Identification and  
RT characterization of the encoded proteins."  
RL J. Biol. Chem. 273:15395-15403(1998).  
DR EMBL; AF055666; AAC27741.1; -.  
DR PFAM; PF00409; kinesin2; 4.  
DR PRINTS; PR00381; KINESINLIGHT.  
SQ SEQUENCE 599 AA; 66662 MW; 0B012E26 CRC32;

Query Match 65.1%; Score 2056.5; DB 11; Length 599;  
Best Local Similarity 69.6%; Pred. No. 1.3e-119;  
Matches 422; Conservative 61; Mismatches 102; Indels 21; Gaps 8;

QY	1	MSGLV	QORDE	PAGHRL	SQEI	LGS	TRVLS	QGLE	ALRSE	HQA	VLQ	SL	SQTIE	CLQGG	HE	60
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	1	MATMVL	-----	PREEKLS	QDEIVL	GTKAVI	QGLE	TLRGE	HALLA	PLAS	-----	HEAGEA				50
QY	61	EGLVHE	KARQL	RRSMENI	ELGL	SEAQ	VMLAL	ASHL	STVE	SEKQ	KLRAQ	VRR	LCQEN	QWLR		120
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	51	EPGSQ	ERCLL	RRSLEA	IELGL	GAQV	ILAL	SSHL	GAVE	SEKQ	KLRAQ	VRR	LQEN	QWLR		110
QY	121	DELAGT	QORL	QRSEQA	VAQLE	EEK	KHLE	FLG	QLRQ	YDE	DGHT	SEEK	EG	DATA	KDSLD	180
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	111	EELAGT	QOKL	QRSEQA	VAQLE	EEKQ	HL	FLF	MSQIR	KLDE	--MLP	QEEK	G	VDVP	KDSLD	168



Db 216 LNIALVYRDQNKYKDAANLLNDALAIAREKTLGKDHPPAATAATLNNLAVLYGKRGKYKEAE 275

QY 316 PLCQRALEIREKVLGTNHPDPAKQNLNALLCQNGQKYEAVERYQORALAIYEGQLGPDN 375

Db 276 PLCKRALEIREKVLGKDHDPDAKQNLNALLCQNGQKYEEVEYYQRALEIYQTKLGPDD 335

QY 376 PNVARTKNNLASCYLKQGYKAEATLYKEILTRAHVQEFSGVDDHKKPIWMHAEREEMS 435

Db 336 PNVAKTKNNLASCYLKQGYKAEATLYKEILTRAHERFSGVDDENKPIWMHAERECK 395

QY 436 KSRHHEGTPYAEYGGWKACKVSSPTVNTLRNLGALYRRQGLKLEAAETLEECALRSRR 495

Db 396 GKQ--KDGTSFGYGGWKACKVDSPTVNTLRNLGALYRRQGFEEAETLEEAAMRSRK 453

QY 496 QGTDPISTKVAELGESDGRRTSQ 520

Db 454 QGLDNVHKQORVAEVLNDPESIEKSE 478

RESULT 4

Q21592 ID Q21592 PRELIMINARY; PRT; 521 AA.

AC Q21592;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE M7.2 PROTEIN.

GN M7.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RA LENNARD N.;

RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., JIER M., JOHNSTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

DR EMBL; 268337; CAA92746.1; -.

DR PFAM; PF00409; kinesin2; 4.

DR PRINTS; PRO0381; KINESINLIGHT.

SQ SEQUENCE 521 AA; 58649 MW; 1BD979DB CRC32;

Query Match 30.4%; Score 962.5; DB 5; Length 521;

Best Local Similarity 44.0%; Pred. No. 4.5e-52;

Matches 220; Conservative 83; Mismatches 166; Indels 31; Gaps 11;

QY 20 EEILGSTRVLSQGLEALRSEHQ---VLQS--LSQTIECQQGGHEEGLVHEKARQLRRS 74

Db 7 DDITTVLKTVOQTLFALRDEHEAATRIEANLSKTIYGMNSS--EPLSPSEKMLIDES 64

QY 75 MENIELGLSEAQVMLALASHLSVSESEKQKRLAQVRRLCQENQWLRLDELACTOQLORSE 134

Db 65 LGKVMGDEASLLIMDKLMQSYDVQLSKNHESIRLLRQENTWLLDELTTTQRLQESE 124

QY 135 QAVAQLEEKHLEFLGQLRQYDED-GHTSEEKGDATKDSLDDLPNNEEEDPSNGLSR 193

Db 125 RTVAHLEERDHYKFDQSMYLNLSDFQHTTSVDATPMVMVDTLQELGFGPEEDONNN-QA 183

QY 194 GOGATAAQ-----QGYEIPARLRTLNLNLVQYAAQGRYEAVPLCKQALEDLERTSGRG 248

Db 184 DOGCRSSFSNPISNDYQLPTRLQTLQNLVQYMEQGRFEVAIPCKQALEDDVVKVHGNV 243

QY 249 HPDVATMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPAATAATLNNLAVLYGKR 308

Db 244 HLDVATMLNLVAIVYRNQENFKDAIYLERKALSIRVQCCGENHHSVAATLNNLAIAYGKR 303

QY 309 GKYKEAPLCQRALEIREKVLGTNHPDPAKQNLNALLCQNGQKYEAVERYQORALAIYE 368

Db 304 GKYKESEPLCKRALEIRKNNLLGNPHDPDAKQNLNLGIVTQOLEYEETENYFKOALSIN 363

QY 369 GOLGPDNPVARTKNNLASCYLKQGYKAEATLYKEILTRAHVQEFSGVDDHKKPIWMHA 428

Db 364 RAFFENHONVIKTKNQLASVFLKQGYKQYEAEMYKNILSKVAI-----TGNKPIWRIA 416

QY 429 EEREEMSKSRHHEGTPYAEYGGW-YKACKVSSPTVNTLRNLGALYRRQGLKLEAAETLE 487

Db 417 EDREERQRN-----GIPKVDDESFNVPNTVMDSNMSTIKNLAAVYRKQKEEAAGTLE 471

QY 488 ECALRSRRQ---GTDPISTQ 504

Db 472 E-ALGAKKQINGGADHTNST 490

RESULT 5

O33965 ID O33965 PRELIMINARY; PRT; 490 AA.

AC O33965;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE KINESIN LIGHT CHAIN.

GN KLC.

OS Plectonema boryanum.

OC Bacteria; Cyanobacteria; Oscillatoriales; Plectonema.

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-UTEX 485;

RX MEDLINE; 97355691.

RA CELERIN M., GILPIN A.A., DOSSANTOS G., LAUDENBACH D.E., CLARKE M.W.,

RA BEUSHAUSEN S.;

RT "Kinesin light chain in a eubacterium.";

RL DNA Cell Biol. 16:787-795(1997).

DR EMBL; U78597; AAB87735.1; -.

DR PFAM; PF00409; kinesin2; 5.

SQ SEQUENCE 490 AA; 55967 MW; 4A7472E2 CRC32;

Query Match 15.9%; Score 503.5; DB 2; Length 490;

Best Local Similarity 33.9%; Pred. No. 9.7e-24;

Matches 164; Conservative 62; Mismatches 153; Indels 105; Gaps 17;

QY 130 LORSEQAVQAQLEEEKHLEFLGQLRQYDEDGHTSEEK-----EGDATKDSLDDLPNE 182

Db 73 LKRVDQGMYLHQLLR--EFFVVKREQRADDSLQQRFYQVYIAEAKRVRD----- 121

QY 183 EEEDPSNGLSRGQGATAAQ-QGYEIPAR-----LRTLNLV--IQYAAQGRYEAVPL 233

Db 122 ---EPEKSLIRESTMIAHLQEAERLARPEQALDLATCLNLWLAELYVAQGRYEAEP 178

QY 234 CKQALEDLERTSGRHPDVATMLNIALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPA 293

Db 179 YVRSLSIHEQOLGADHLDVANSPNNLALLYKEQGRYEEAEPLVYRSLSIREQQLGTDHLD 238

QY 294 VAATLNNLAVLYGKRGKYKEAEPICQRALEIREKVLGTNHPDPAKQNLNALLCQNGKY 353

Db 239 VATSLNNLAVLYRSQGRYEEAEPLVYRSLSIREQQLGTDHLDVATSLSLNLAFLYQSGCH 298

QY 354 EAVERYQORALAIYEGQLGPDNPVARTKNNLASCYLKQGYKAEATLYKEILTRAHVQE 413

Db 299 HKAEPILLVRAPIWEQQLGTDHPDVATSLNNLAFLYHLQGRYEEAEPLLVRLS-IREQQ 357

```
QY 414 FGSVDDHKKPIWMHAEEREEMSKSRHHEGGTPYAEYGGWKACKVSSPTVTNTLRLNGAL 473
Db 358 LGT---DH-----PDVATSLNNLAVL 375
QY 474 YRRQGLKEAAETLEECALRSRRQ---GTDPISTQKVAEL--LGESDGRRTSQEG----- 522
Db 376 YHLQGRYEDAEPLLLYSVRIRQEQLPADHPLSAKNLSNLAYLYDLQGRSGEAEALYLQAI 435
QY 523 PGDSVKFEGGEDASVAVIEWSGDSGLQSGSLGKIRDVLR--RSSEL-----LVRK 572
Db 436 PILSAKLEESH-----QWRQEAS---QRFERSL--LQKALQENRTDELSDDPMTQSILQE 484
QY 573 LQGT 576
Db 485 LRST 488

RESULT 6
Q9XBP6 PRELIMINARY; PRT; 1049 AA.
AC Q9XBP6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SERINE/THREONINE KINASE PKN8.
GN PKN8.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DZF1;
RA INOUE S., JAIN R., UEKI T., NARIYA H., XU C., HSU M.,
RA MUNOZ-DORADO J., FAREZ-VIDAL E., INOUE M.;
RT "Sequence Analysis of 13 Eukaryotic-like Protein Ser/Thr Kinases of
RT Myxococcus xanthus, a Developmental Bacterium and Significance of
RT Their Coexistence with Protein His Kinases.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159691; AAD42856.1; -.
KW Kinase.
SQ SEQUENCE 1049 AA; 114313 MW; E93CBA34 CRC32;
```

```
Query Match 8.7%; Score 276.5; DB 2; Length 1049;
Best Local Similarity 25.0%; Pred. No. 2.8e-09;
Matches 158; Conservative 73; Mismatches 243; Indels 157; Gaps 24;

QY 1 MSGVLVQGRDEPAGHRLSQEEILGSTRLV--QGLEALRSEHQAVLQSLSTQTECLQGG 58
Db 502 LEGLV-GCRDTPG-----LTSRPQPPDNLRSRVHAARHKLQAVRAHLVSHR 546
QY 59 HEEGLVHEKARQLRRSMENIELGLSEAQVMLALASHLSTVESEK-----QKLR----- 107
Db 547 FTDGLAVTSA--LLEEQGLGYKPLEAEIFLAHGTVLGGLNKPKEAEAFYQALWAAEAA 604
QY 108 ----QVRLCQENQWLRLDELACTQORLQORSEQAVQAQLEBEKKHLEFLGQLROYDEDGHTS 163
Db 605 RDETVARAWLELIW----AVGEEQ--SRPDEAEKLIRHARAVERLGR----- 647
QY 164 EEKEGATKDSLDDLFNPEEEEDPSNGLSRGQATAAQG--GYEI-----PARLRT-- 213
Db 648 -ERFPDITELHTRLSSLRE-----AYGQIAEAQEQARQGLEFSRRRNHPDSLRTPN 698
QY 214 -LHNLVIQYAAQGRYEVAVPLCKQALEDLERTSGRGHPDVATMLNILALVYRDONKYKEA 272
Db 699 LIHQLGRIFGQGRYEDALKLHREALELRILGADNPALVTSYNRVATASLEVGRYAEA 758
QY 273 AHLNDALSIRESTLGPDPHPAVAATLNNLAVLYGKRKYKEAEPLCQRALEIREKVLGTN 332
Db 759 VSAWRKALALQEAASSDPETTPMGITVLLNLAVSVRVAGRLLEARSMAERARAIFERAGPN 818
QY 333 HPDVAKQLNNLALLCQNOQKYEAVERYYQRALAIYEGQLGPDNPVNARTKNNLASCYLKQ 392
```

```
Db 819 HITVIFALSALADISSEAGQGDALGLATEALERVORSIGQDTPRAALPLTVRGQVYLKA 878
QY 393 GKYAEAEATLYKEILTRAHVQEFSGVDDHKKPIWMHAEEREEMSKSRHHEGG-----TPY 446
Db 879 GRHAEARDDLALQOR-----LEKEHGP-----EGGKTVTVLLPL 913
QY 447 AEYGGWKACKVSSPTVTNTLNLGALYRRQCKLEAAETLEECALRSRRQGTDPISQTKV 506
Db 914 AEF-----ALATRASK-----EALAYCE-RAR-----KV 936
QY 507 AELLGESDGRRTSQEGPGDSVKFEGGEDASVAVIEWSGDSGLQSGSL-----GKIRD-- 560
Db 937 TEKAG-----GAESPDGASALACSGEAHLALGTAAEAVPLLERARRIQTQNGEVKDPR 989
QY 561 VLRRSELLVRKLGQTEPRPSSSNMKRAASL 591
Db 990 VAGKTAFLARALMETRASP-----ERARAL 1015

RESULT 7
O81629 PRELIMINARY; PRT; 609 AA.
ID O81629;
AC O81629;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE F8M12.21 PROTEIN.
GN F8M12.21.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA WASHU;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA MADSEN C., GRAVES T., COTTON M., MODDE T.;
RT "The sequence of A. thaliana F8M12.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA WATERSTON R.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF080118; AAC33943.1; -.
DR PFAM; PF00515; TPR; 2.
SQ SEQUENCE 609 AA; 66280 MW; D071143A CRC32;
```

```
Query Match 7.3%; Score 232; DB 10; Length 609;
Best Local Similarity 21.9%; Pred. No. 7.7e-07;
Matches 106; Conservative 66; Mismatches 183; Indels 130; Gaps 13;

QY 213 TLHNLVIQYAAQGRYEVAVPLCKQALEDLERTSG-----RGHPDVATMLNILALVYR 264
Db 142 SLHVLAAYICSLGRFDEAVPPLERAIQVPDPTRGPDHSLAAFSGHMQLGDTLSMLGQIDR 201
QY 265 DQNKYKEAAHLLNDALSIRESTLGPDPHPAVAATLNNLAVLYGKRKYKEAEPLCQRALEI 324
Db 202 SIACYEE-----GLKIQIQTIGDTPRVGETCRYLAEAYVQAMQFNKAEELCKKTLEI 254
QY 325 REKVLGTNHPDPAKQLNNLALLCQNOQKYE-AVERY----- 359
Db 255 HRAHSEPASLEEAADRRMLAICEAKGDYENALEHLVLASMAMIASQGESEVASIDVSI 314
QY 360 -----YQRALAIYEGQLGPDNPVNARTKNNLASCYLKQKYAEAE----- 400
```



Db 315 NIYMSLCRFDEAVFSYQKALTVPKASKGETHTPTVASVVRVLAELYHRTGKLBESKSYCEN 374  
QY 401 ---LYKEILLTRAHVQEFSGVDDDDHKPIWMHAAEEREEMSKSRHHEGGTPYAEYGGWYKACK 457  
Db 375 ALRIYNKVPVPGTTVEEIIAGGLTEISAIYESVDEPEEALKLLQ-----KSMK 420  
QY 458 V--SSPTVNTTLRNL---GALYRRQKLEAAETLEECALRRRQGTDP-----ISQT 504  
Db 421 LLEDKPGQSAIAGLEARMGVYTVGRYEDARNAFESAATKLRAAGEKSAFFGVVNLQM 480  
QY 505 KVA-----ELLGESDGRRTSQEGPDSVKFEGGEDASVAVEMSGDGSGLQSRG 553  
Db 481 GLACVQLFKIDEAGELFEEARGILEQERGRCDQ-----DTLGVYSNLAATYD 527  
QY 554 SLGKIRD---VLRSSSELLVRKLGQTEP-----RPSSSNMKRAASLNYLNQ 596  
Db 528 AMGRIEDAIEILEQVLKREKLGITANPDFEDEKKRIAEALLKEAGRSRNYKAKSLQNLID 587  
QY 597 PSAAP 601  
Db 588 PNARP 592

RESULT 8  
O86768 ID O86768 PRELIMINARY; PRT; 886 AA.  
AC O86768;  
DT 01-NOV-1998 (TremBLrel. 08, Created)  
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)  
DE PUTATIVE ATP /GTP-BINDING PROTEIN.  
GN SC6A9.38.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA MURPHY L., HARRIS D.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE; 97000351.  
RA REDENBACH M., KIESER H.M., DENAPAITTE D., EICHNER A., CULLUM J.,  
RA KINASHI H., HOPWOOD D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL031035; CAA19923.1;  
SQ SEQUENCE 886 AA; 97715 MW; 848BB121 CRC32;

Query Match 7.3%; Score 229.5; DB 2; Length 886;  
Best Local Similarity 23.5%; Pred. NO. 1.8e-06;  
Matches 113; Conservative 74; Mismatches 216; Indels 77; Gaps 14;

QY 1 MSLGLVQRDEPAGHRLSQEEILGSTRL-----VSQGLEALRSEHQAVLSLSQTIE 52  
Db 341 LQGRVGMDEQADTLAAD---LGDPLALAAQAAGVIANGVSLGLYHQLL---ATNTTR 394  
QY 53 CLQQGGHEEGLVHEKARQLRRSMENIELGLSEAQVLMALASHL-----STVESEKQL 105  
Db 395 ILERGG-APGYPAPVAATVTIADVRLTDEHPDAIALRLVALLGPEPIRNAWLESVRPRL 453  
QY 106 RAQVRLCQENQWLDELQAGTQQLRQSRSEQAVALEEKKHLEFLGQ---LR----- 154  
Db 454 -TTIPGSDDDPMWLHEALS-----PLSRYGLARTEPESFQIHRLTQAVLRDQSLPDQSA 506

QY 155 --QYDEDGHTSEEKGDATKD-----SLDDLFPNEEEE-----DPSNGLSR 193  
Db 507 AIRRDATAALLSAANPGDPQSPGNWPGWAALTPLHTAQHLAPTEQPELRPTLLDAAHYLIR 566  
QY 194 -GQGATA-----AQGGYEIPARLRTLHNLVYQAAQGRYEVAVPLCKQALEDL 241  
Db 567 SGQTRTARDLTATAHAAWATDLGEDHPDTLTSAQFLGHATADLGEHAEARRIETDLTRR 626  
QY 242 ERTSGRHPDVATMLNILALVYRDQNKYKEAAHLLNDALSIRESTLGPDPHPAATLNNL 301  
Db 627 RRTLGDDHPHTLQSANDLANVLDNLGEHTESHRIETLRRRRTLGGDDHPHTLQSAHNL 686  
QY 302 AVLYKRGKYKEAEPLCQRALEIREKVLGTNHPDVAKQLNNLALLCQNOQKYEAVERYQ 361  
Db 687 AASLHHLGEHTEARRILEDTLRRRTLGDDHLDLTLSAHSIAASLHSLGEHTESVRMDQ 746  
QY 362 RALAIYEGQLGPDNPNVARTKNNLASCYLKQGYAEATLYKEILTRAHVQEFSGVDDDDH 421  
Db 747 ETLARLRLILGDDHPDTLQSAHNLAVTLHDLGEHTEARRIETLTR---QRRTLGDDH 802

RESULT 9  
Q92W40 ID Q92W40 PRELIMINARY; PRT; 707 AA.  
AC Q92W40;  
DT 01-MAY-1999 (TremBLrel. 10, Created)  
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TremBLrel. 10, Last annotation update)  
DE F17L21.30.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,  
RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,  
RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSIEL N.A., THEOLOGIS A.,  
RA ECKER J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC F17L21."  
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AC004557; AAC80630.1;  
SQ SEQUENCE 707 AA; 77828 MW; 89334300 CRC32;

Query Match 6.2%; Score 196.5; DB 10; Length 707;  
Best Local Similarity 21.6%; Pred. NO. 0.00015;  
Matches 123; Conservative 85; Mismatches 226; Indels 135; Gaps 22;

QY 10 DEPAGHRLSQEEIL--GSTRLVSGQ--LEALRSEHQAVLSLSQTIECQQGG----- 58  
Db 194 EEPSENVELARFLNQARNLVSSGSDTHKALELTHRA-----AKLFEASAENGKPCLEW 248  
QY 59 ----HEEGLVHEKARQLRRSMENIELGLSEAQVLMALASHLSTVESEKQLRAQVRLCQ 114  
Db 249 IMCLHVTAAVHCKLKE-----YNEAIPVLQRSVEIPVVEEGEHALAKFAGLMQ 297  
QY 115 ENQWLDELQAGTQQLRSEQAVALLEE---EKKHL-----EFLGQLRQ 155  
Db 298 -----LGDTYAMVGQLESSISCYTEGLNIQKVLGENDPRVGETCRYLAELVQALR 349  
QY 156 YDEGHTSEEKGDATKDSLDLFPNEEEEDPSNGLSRGQATAAQGGYEIPARLRTLH 215  
Db 350 FDEAQQVCETA-----LSIHRESGLP-----GSIAEAD-----RRLM 382  
QY 216 NLVYQAAQGRYEVAVPLCKQALEDLERTSGRHPDVATMLNILALVYRDQNKYKEAAHL 275  
Db 383 GLICE--TKGDHENA--LEHLVLASMAAMAANGQSEVAFVDTSIGDSYLSLRFDEAICA 438  
QY 276 LNDALSIRESTLGPDPHPAATLNNLAVLYKRGKYKEAEPLCORALEIREK-VLGTNHP 334

Db 439 YOKSLTAKTAKGENHPAVGSVYIRLADLYNRTGKVREAKSYCENALRIYESHNLISPE 498

QY 335 DVAKQLNNLALLCONQKYEAVERYYQRALAIYEGQLGPDNPN-----VARTKNNLASCYL 390

Db 499 EIASGLTDSIVICESMNEVEQAITLLQKALKIY-----ADSPGQKINAGIEAQMGVLYY 553

QY 391 KQKGYAEATLYKEILTRAH-----VQFSGVDDHDPFIMWHAEREEMSKS 437

Db 554 MMGKYMESYNTFKSAISKLRATGKKQSTFFGIALNQMLACIQLDATFEAEVLFEEAKCI 613

QY 438 RHHEGTPYAEYGGWKACKVSSPTVNTTLRLNLAGLYRRQKLEAAETL--EECALRSRR 495

Db 614 LEQECGPYHPETGLYS-----NLGAYDAIGRLDDAIKLLGHVVGVRREEK 659

QY 496 QGT-DPISQ---TKVAELL--GESDGR 517

Db 660 LGTANPVTEDKRRRLAQLLKEAGNVTGRK 688

RESULT 10

O67735

ID O67735 PRELIMINARY; PRT; 342 AA.

AC O67735;

DT 01-AUG-1998 (TReMBLrel. 07, Created)

DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)

DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)

DE HYPOTHETICAL 40.4 KD PROTEIN.

GN AQ.1896.

OS Aquifex aeolicus.

OC Bacteria; Aquificales; Aquificaceae; Aquifex.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VF5;

RX MEDLINE; 98196666.

RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L., GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R., RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;

RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.";

RL Nature 392:353-358(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-VF5;

RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L., GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R., RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE000762; AAC07708.1; -.

DR PFAM; PF00515; TPR; 3.

KW Hypothetical protein.

SQ SEQUENCE 342 AA; 40375 MW; 58842C97 CRC32;

Query Match 6.0%; Score 189; DB 2; Length 342;

Best Local Similarity 25.1%; Pred. No. 0.00016;

Matches 88; Conservative 37; Mismatches 119; Indels 106; Gaps 13;

QY 221 YAAQGRYEVAVPLCKQALDERTSGRGRHPDVATMLNILALVYRDQNKYKEAAHLLNDAL 280

Db 36 FAAR-NYSEAIANFHKAL-----RANPDEPRIWALGAYLMEAKYKAEESFKKAL 86

QY 281 SIRESTLGPDPHVAATLNNLAVLYGKRGKYEAPLCQRA----- 321

Db 87 SI-----NPNYSEARKNLGILYKLGRYEALKYLQEAANDEYVEKKHEAFYLLAKV 138

QY 322 -----LEIREKVLGTNHPDVAKQLNNLALLCONQKYEAVERYYQRALAIYEGQL 371

Db 139 YEAKQDLKNVRYLEKAVAYNPNFVQAQL-ELAAQAYENLGKYEAEKIYKSLLL----- 191

QY 372 GPDNPNVARTKNNLASCYLKQKGYAEA-----ETLYKEILT---RAHVQEEFGS----- 416

Db 192 --NGFNKPFLLKLAEVYKKG DYERAREITIKELLYKENLTNEQREKVELLTKVLLAQ 249

QY 417 ----VDDDHKPIWMHAEEREEMSKSRHHGGTPYAEYGGWKACKVSSPTVNTTLRLNLA 472

Db 250 RKLIIPRVHKPI-----KKEKKKEKY-----YAV-----QLGA 278

QY 473 LYRRQKLEAAETLEECALRSRR-QGTDPISQTKVAELLGESDGRRTSQE 521

Db 279 FSTKERADKLVOELKSKGLRDLRLPDTGVYKVIYGRFETPEARRAKEE 328

RESULT 11

Q55489

ID Q55489 PRELIMINARY; PRT; 948 AA.

AC Q55489;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)

DE HYPOTHETICAL 104.9 KD PROTEIN.

OS Synechocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PCC6803;

RA TABATA S.;

RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-PCC6803;

RX MEDLINE; 96127529.

RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N., RA SUGIURA M., TABATA S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome.";

RL DNA Res. 2:153-166(1995).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-PCC6803;

RX MEDLINE; 97061201.

RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y., RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T., RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S., RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M., RA TABATA S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";

RL DNA Res. 3:109-136(1996).

DR EMBL; D64006; BAA10838.1; -.

DR PFAM; PF00515; TPR; 5.

KW Hypothetical protein.

SQ SEQUENCE 948 AA; 104851 MW; 698807BB CRC32;

Query Match 6.0%; Score 189; DB 2; Length 948;

Best Local Similarity 22.1%; Pred. No. 0.00063;

Matches 118; Conservative 74; Mismatches 176; Indels 166; Gaps 22;

QY 2 SGLVLGQDEPAGHRL----SQEILGSTRLVSQGLEALRSEHQAVLQSLSQTIACLQOG 57

Db 19 AGILINQID-PQGYGVVVTSNKSGVTGKYKL-----QIILPDANQLVEI---Q 62

QY 58 GHEEGLVHEKARQLRRSMENI-----ELGLSEAQVMLALASHLSTVES-- 100

Db 63 GH-QSLIKAQAMMDEGSKESLLOAIELFTQGLEFYRQIGIVNRETFFAL-NRLGTIYSDF 120

QY 101 -EKQKLRAQVRRRLCQENQWLDE-LAGTQORLORSEQAVAQLEEEKKHLEFLGQ----LR 154

Db 121 GEKSQALAYYEQAIPLAQOQDKALEGA--TLNNGISYINALADRKKAIDFYQQAALVLR 178

QY 155 QYDEDGHTSEEKGGATKDSLDDLFPNNEEDPSNGLSRGQGATAAQGGGYEIPARLRTL 214

Db 179 Q-----AGDKTEE-----MTTI 190





QY 338 KOLNNLALLCONOGKYEAVERYYQRALAIYEGQLGPDNPNVARTKNNLASCYLKQG--KY 395  
Db 1845 EEL-----QKAYKYESIKR-----TRLTYDKEEL-QYHLKQRSIQL 1879  
QY 396 AEATLYKEILTRAHVQEFSGVDDDDHKPIWMHAEEREEMSKSRHHEGGTPYAEGGWYKA 455  
Db 1880 QSAESKLQDLSTGSHDNSLSS-----HSRCSLGRS-----GLEIA 1914  
QY 456 CKVSSPT-----VNTTLR-----NLGALYRROGKLEAAETLEECALRSR 494  
Db 1915 VTTSSPTSPYKGMIERNDVSVWTLIEDESFKGTQLRLLRAGSLRSNN--ERCPIQRR 1972  
QY 495 RQGT-----DPISQTKVAELL-----GESDGRRTSQEGPGDSVK 528  
Db 1973 QTLTNSGHNSGAINGGSSPAHPNPLSQSMSATALLRSSNSGSESEGRPLSR----- 2024  
QY 529 FEGGEDASVAVEWSGDGTQLRSGSLGKIRDVLR-----RSELLVRLKQGT 577  
Db 2025 ---ARSKSMCIKASATSSAVCESSGQKKPKQDELSLADWPEDIPLCSSP----QAPGIE 2077  
QY 578 PRPSSNMK-----RAASLNLYNQPSAAPLOVS-----RGLSASTMDLSSSS 619  
Db 2078 MRPRSSTMKLSSEAKKFQEI-QESAGEAMVSGANSEDESCASSEDIMRSS 2128  
RESULT 14  
O02244  
ID O02244 PRELIMINARY; PRT; 1963 AA.  
AC O02244;  
DT 01-JUL-1997 (TremBLrel. 04, Created)  
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)  
DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)  
DE UNC-54 PROTEIN.  
GN UNC-54.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KERSHAW J.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL; 283107; CAB05505.1; -.  
DR EMBL; 281499; CAB05505.1; JOINED.  
DR EMBL; 281499; CAB04089.1; -.  
DR EMBL; 283107; CAB04089.1; JOINED.  
DR HSSP; P13538; 2MYS.  
DR PFAM; PF00063; myosin\_head; 1.  
DR PFAM; PF01576; Myosin\_tail; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
SQ SEQUENCE 1963 AA; 224754 MW; DACC0BF9 CRC32;

Query Match 5.7%; Score 179; DB 5; Length 1963;  
Best Local Similarity 22.1%; Pred. No. 0.0068;  
Matches 158; Conservative 120; Mismatches 254; Indels 184; Gaps 36;

QY 30 SQGLEALRSEHQAVLQSLQSTIECLOQGGHEEG--LVHEKARQLRRSMENIELGLSEAQV 87  
Db 855 AEELEKINDVKALEDSLAK-EXLRKELEEESAKLVEEXT-SLFTNLESTKTQLSDAEE 912  
QY 88 MLA-----LASH---LSTVESEKQKILRAQV-----RR 111  
Db 913 RLAKLEAQOKDASKQLSELNDQLADNEDRTADVQRAKKKTEAEVEALKKQIQDLEMSLRK 972  
QY 112 LCOENQWLDELQAGTQORLQORSEQAVALLEEKKHLEFLQLRQYDEDEGHTSEEKGDAT 171  
Db 973 AESEKQSKDHQIRSLQDEMQQQDEAIKLNKKEKKHQQEIN--RKLMEDLQSEEDKGNHQN 1030  
QY 172 K-----DSLDDLFPEEEEE-----DPSNGLS 192  
Db 1031 KVKAKLEQTLDDLEDSLEREKARADLDKQKRVGELKIAQENIDESGRQRHDLNNLK 1090  
QY 193 RGQATAAQGGYEIPARLRTLHNLVIOYAAQ---GRYEVAVPLCKQALEDLER----- 243  
Db 1091 KKESEL-----HSVSSRLEDEQALVSKLQRIKDGQSRIS-----ELEEELENERQSR 1139  
QY 244 TSGRHGPDVATMLNILALVYRQ-----NRYKEA--AHL---LNDALSIRESTLG 288  
Db 1140 KADRAKSDLQRELEELGKLEGGGATAAQVEVKKREAEAKLRRDLEAANNHNLQ 1199  
QY 289 ----PDHPAVAATLNNLAVLYGKRGK-----YKEAEPLCQRALEIREKVLG-TNHPD 335  
Db 1200 GLRKKHTDAVAELTDQLDLNKAQAKVEKDKAQAVRDAEDL---AAQLDQETSGKLNNEK 1256  
QY 336 VAKQLNNLALLCQNGKYEAVERYYQRALAIYEGQLGPDNPNVARTKNNLASCYLKQGY 395  
Db 1257 LAKQFE--LQTELQSKADEQSRQLQDFTSL-KGRHSENGDLVR-----QL 1300  
QY 396 AEATLYKEILTRAHVQEFSGVDDDDHKPIWMHAEEREEM---SKSRHHEG-----GTPYAE 448  
Db 1301 EDAESQVQ-N-LTRLKSQLTSOLEEARRTADEARERQTVAAQAKNYQHEAEQLQESLEE 1359  
QY 449 YGGWYKACKVSSPTVNTTLNLGALYRROGKLEAAETLEECALRSRRQGTDPISQTKVAE 508  
Db 1360 IEGKNEILRQLS-KANADIQOMKARFEGEGLLKADE-LED---AKRRQAQKINE--LQE 1411  
QY 509 LLGESDGRRTSQEGP-----GDSVKFEGGEDASVAVE-WSGDGTQLRSGSLGKIRDVL 562  
Db 1412 ALDAANSKNASLEKTKSRVLVGD-----DDAQDVERANGVASALEKKQKGFDKIIDW 1465  
QY 563 RRSELLVRKLOGTEP--RPSSNMKRAASLNLYNQPSAAPLOVSRGLSASTMDLS 616  
Db 1466 RKKTDDLAELDGAQRDLRNTSTDLFKAKNA-----QEELA--EVVEGLRRENKSL 1515  
RESULT 15  
Q9Y148  
ID Q9Y148 PRELIMINARY; PRT; 1011 AA.  
AC Q9Y148;  
DT 01-NOV-1999 (TremBLrel. 12, Created)  
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)  
DE BCDNA.GH04245 (FRAGMENT).  
GN BCDNA.GH04245.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA RUBIN G.M., WAN K.H., HARVEY D., LEWIS S.E., BROKSTEIN P., TSANG G.,  
RA AGBAYANI A., ARCAINA T.T., BAXTER E., BLAZEJ R.G., BUTENHOFF C.,  
RA CHAMPE M., CHAVEZ C., CHEW M., DOYLE C.M., FARFAN D.E., FRISE E.,  
RA GALLE R., GEORGE R.A., HARRIS N.L., HOSKINS R.A., EVANS-HOLM M.,  
RA HOUSTON K.A., HUMMASTI S.R., KIM E., LI P., MOSHREFI M., PACLEB J.M.,  
RA PARK S., SEQUEIRA A., SETHI H., SNIR E., SVIRSKAS R.R., WEINBURG T.,  
RA CELNIKER S.E.;  
RT "Full length Drosophila melanogaster cDNA sequence.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.



Query Match	5.6%;	Score	178;	DB 5;	Length	1011;
Best Local Similarity	24.8%;	Pred. No.	0.0033;			
Matches 110;	Conservative	56;	Mismatches	187;	Indels	90;
					Gaps	18;

Search completed: August 15, 2000, 13:17:15  
Job time: 9045 sec



GenCore.version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2000, 00:39:17 ; Search time 2177.55 Seconds  
(without alignments)  
2010.452 Million cell updates/sec

Title: US-09-036-614A-2  
Perfect score: 2453  
Sequence: 1 GTGAAGTGGTGAACAAGGG.....AGAATCAGTCTTNTNTNNG 2453

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_bal:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: em\_fun:\*  
17: em\_hum1:\*  
18: em\_hum2:\*  
19: em\_in:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_sy:\*  
29: em\_un:\*  
30: em\_vi:\*  
31: gb\_htg1:\*  
32: gb\_htg2:\*  
33: gb\_in1:\*  
34: gb\_in2:\*  
35: em\_bal:\*  
36: em\_ba2:\*  
37: em\_hum3:\*  
38: em\_hum4:\*  
39: gb\_pr4:\*  
40: gb\_htg3:\*  
41: gb\_htg4:\*  
42: gb\_htg5:\*  
43: gb\_htg6:\*

44: gb\_htg7:\*  
45: em\_htg1:\*  
46: em\_htg2:\*  
47: em\_htg3:\*  
48: em\_hum5:\*  
49: gb\_pl3:\*  
50: gb\_pr5:\*  
51: gb\_htg8:\*  
52: gb\_htg9:\*  
53: gb\_htg10:\*  
54: gb\_htg11:\*  
55: gb\_htg12:\*  
56: gb\_htg13:\*  
57: gb\_htg14:\*  
58: gb\_in3:\*  
59: gb\_htg15:\*  
60: gb\_htg16:\*  
61: gb\_htg17:\*  
62: em\_htg4:\*  
63: em\_htg5:\*  
64: em\_htg6:\*  
65: em\_htg7:\*  
66: em\_hum6:\*  
67: gb\_htg18:\*  
68: gb\_htg19:\*  
69: gb\_htg20:\*  
70: gb\_htg21:\*  
71: gb\_htg22:\*  
72: gb\_htg23:\*  
73: gb\_htg24:\*  
74: gb\_htg25:\*  
75: gb\_htg26:\*  
76: gb\_htg27:\*  
77: gb\_htg28:\*  
78: gb\_htg29:\*  
79: gb\_htg30:\*  
80: gb\_htg31:\*  
81: gb\_vil:\*  
82: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1479.4	60.3	1524	11 HSM802612	AL162078 Homo sapi
2	841.2	34.3	2840	12 AF055666	AF055666 Mus muscu
3	785.4	32.0	1917	12 RATKINLB	M75147 Rat kinesin
4	785.4	32.0	2386	12 RATKINLC	M75148 Rat kinesin
5	783.8	32.0	2308	12 RATKINLA	M75146 Rat kinesin
6	762.8	31.1	2055	12 CGKINID	Y14586 Cricetulus
7	721.2	29.4	2289	12 AF055665	AF055665 Mus muscu
8	713.4	29.1	2308	9 HUMKINESLC	L04733 Homo sapien
9	593.4	24.2	2601	33 SUSKINLCIA	L10233 Strongyloce
10	593.4	24.2	2685	33 SUSKINLCIB	L10234 Strongyloce
11	593.4	24.2	2712	33 SUSKINLCIC	L10235 Strongyloce
12	571.8	23.3	1846	33 DROLCKIN	L11013 Drosophila
13	571.8	23.3	1846	33 DROLCKIN	L11328 Drosophila
14	560.6	22.9	9915	14 AF055298	AF055298 Expressio
15	540.8	22.0	2043	33 SQDSKLC	L24440 Loligo peal
16	532.8	21.7	2922	33 SUSKINLCID	L08258 Strongyloce
17	507.4	20.7	3089	4 GGU48359	U48359 Gallus gall
18	483	19.7	1930	33 SQDSKLCIA	L24441 Loligo peal
19	404.4	16.5	2035	33 CEKINLC1	Z29644 C.elegans m
20	403.8	16.5	1947	33 CEKINLC2	Z29645 C.elegans m
c 21	332	13.5	43154	34 CELC18C4	U55369 Caenorhabdi
22	299.6	12.2	95961	42 AC017329	AC017329 Drosophil
23	299.6	12.2	297385	34 AE003540	AE003540 Drosophil
c 24	291.2	11.9	90514	32 AL136304	AL136304 Homo sapi

```
25 224.2 9.1 23470 33 CEM7
26 224.2 9.1 84118 31 CEY5F2
27 160.4 6.5 1844 2 PBU78597
c 28 148.2 6.0 16440 51 AC023677
29 146 6.0 64412 31 AP000630
30 146 6.0 164556 31 AP000759
c 31 146 6.0 185461 31 AP001107
32 145 5.9 186908 31 AP001812
c 33 142.4 5.8 4623 12 CRUKLCG1
c 34 140 5.7 32595 12 MUSXPDG1
c 35 139.8 5.7 54336 11 HUMXPDG1
c 36 139.8 5.7 160069 54 AC023329
c 37 139.2 5.7 205054 40 CNS01DXG
38 120 4.9 85780 56 AC023803
39 77.8 3.2 660 8 CNS01D3X
40 77.8 3.2 1150 81 HS4ULIR3
41 77.8 3.2 5452 14 U02454
42 77.8 3.2 9600 5 A92665
43 77.8 3.2 10596 5 I25041
44 77.8 3.2 10596 5 I30503
45 77.8 3.2 10737 14 XXU02428
```

```
268337 Caenorhabdi
298871 Caenorhabdi
U78597 Plectonema
AC023677 Drosophil
AP000630 Homo sapi
AP000759 Homo sapi
AP001107 Homo sapi
AP001812 Homo sapi
L47236 Cricetulus
L47235 Mus musculu
L47234 Homo sapien
AC023329 Homo sapi
AL139300 Homo sapi
AC023803 Mus muscu
AL116341 Botrytis
J02079 epstein-bar
U02454 Cloning vec
A92665 Sequence 1
I25041 Sequence 15
I30503 Sequence 15
U02428 Cloning vec
```

```
polyA_signal
polyA_site 382 a 421 c 436 g 285 t
BASE COUNT
ORIGIN
```

```
Query Match 60.3%; Score 1479.4; DB 11; Length 1524;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1480; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 961 GACCAGAATAAGTATAAGGAAGCTGCCACCTGCTGAATGATGCCCTTAGCATCCGGGAG 1020
Db 1 GACCAGAATAAGTATAAGGAAGCTGCCACCTGCTGAATGATGCCCTTAGCATCCGGGAG 60
QY 1021 AGCACCTTGGGACCTGACCATCCTGCTGTGGCTGCCACACTCAACAATTTGGCTGTGCTC 1080
Db 61 AGCACCTTGGGACCTGACCATCCTGCTGTGGCTGCCACACTCAACAATTTGGCTGTGCTC 120
QY 1081 TATGGCAAAAGGGGCAAGTACAAGGAGGCAGAGCCTCTGTGCCAGCGGCACTGGAGATT 1140
Db 121 TATGGCAAAAGGGGCAAGTACAAGGAGGCAGAGCCTCTGTGCCAGCGGCACTGGAGATT 180
QY 1141 CGAGAAAAGGTCCTGGGCACGAATCATCCAGATGTGGCAAAAACAGCTGAACAACCTTGCC 1200
Db 181 CGAGAAAAGGTCCTGGGCACGAATCATCCAGATGTGGCAAAAACAGCTGAACAACCTTGCC 240
QY 1201 CTCTTGTGCCAAAACCCAGGGCAAGTATGAGGCCCTGGAACGCTACTACCAGCGAGCACTG 1260
Db 241 CTCTTGTGCCAAAACCCAGGGCAAGTATGAGGCCCTGGAACGCTACTACCAGCGAGCACTG 300
QY 1261 GCCATCTACGAGGGGCAGCTGGGGCCGGACACACCCCTAATGTAGCCCCGGACCAAGAACAAC 1320
Db 301 GCCATCTACGAGGGGCAGCTGGGGCCGGACACACCCCTAATGTAGCCCCGGACCAAGAACAAC 360
QY 1321 CTGGCTTCCTGTTACCTGAAACAGGGCAATATCTGAGGCTGAGACACTATACAAGAG 1380
Db 361 CTGGCTTCCTGTTACCTGAAACAGGGCAATATCTGAGGCTGAGACACTATACAAGAG 420
QY 1381 ATCCTGACCCCGTCCCATGTACAGGAGTTTGGTCTGTGGATGTATGACCAAGCCCATC 1440
Db 421 ATCCTGACCCCGTCCCATGTACAGGAGTTTGGTCTGTGGATGTATGACCAAGCCCATC 480
QY 1441 TGGATGCATGCAGAGGCGGAGGAAATGAGCAAAAGCCGGCACCATGAGGGTGGGACA 1500
Db 481 TGGATGCATGCAGAGGCGGAGGAAATGAGCAAAAGCCGGCACCATGAGGGTGGGACA 540
QY 1501 CCTATGCTGAGTATGAGGCTGGTACAAAGGCTGCAAAGTGAGCAGCCCCACAGTGAAC 1560
Db 541 CCTATGCTGAGTATGAGGCTGGTACAAAGGCTGCAAAGTGAGCAGCCCCACAGTGAAC 600
QY 1561 ACTACTCTGAGAAACCTGGGAGCTCTGTATAGGCCGCCAGGAAAGCTGGAGGCTGCTGAG 1620
Db 601 ACTACTCTGAGAAACCTGGGAGCTCTGTATAGGCCGCCAGGAAAGCTGGAGGCTGCTGAG 660
QY 1621 ACCCTGGAGGAATGTGCCCTGCGGTCCCGGAGACAGGGCACTGACCCTATCAGCCAGACG 1680
Db 661 ACCCTGGAGGAATGTGCCCTGCGGTCCCGGAGACAGGGCACTGACCCTATCAGCCAGACG 720
QY 1681 AAGGTGGCAGAGCTGCTTGGGAGAGTATGATGTAAGGACCTCCAGGAGGGCCCTGGA 1740
Db 721 AAGGTGGCAGAGCTGCTTGGGAGAGTATGATGTAAGGACCTCCAGGAGGGCCCTGGA 780
QY 1741 GACAGTGTGAAATTCGAGGGTGGTGAAGATGCTTCTGTGGCTGTGGAGTGGTCCGGGGAT 1800
Db 781 GACAGTGTGAAATTCGAGGGTGGTGAAGATGCTTCTGTGGCTGTGGAGTGGTCCGGGGAT 840
QY 1801 GGCAGTGGGACCCCTGCAGAGGAGTGGCTCTCTTGGCAAGATCCGGGATGTGCTCCGCAGA 1860
```

ALIGNMENTS

```
RESULT 1
HSM802612 1524 bp mRNA 23-MAR-2000
LOCUS HOMO sapiens mRNA; CDNA DKFZp761H229 (from clone DKFZp761H229);
DEFINITION partial cds.
ACCESSION AL162078.1 GI:7328159
VERSION .
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1524)
JOURNAL Blum,H., Bauersachs,S., Mewes,H.W., Weil,B. and Wiemann,S.
COMMENT Submitted (15-MAR-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp761H229) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
Location/Qualifiers
1..1524
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761H229"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="amygdala"
1..1068
/gene="DKFZp761H229"
<1..1068
/gene="DKFZp761H229"
/note="similarity to kinesin light chain (Homo sapiens)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB82411.1"
/db_xref="GI:7328160"
/translation="DQNKYKAAHLLNDALSIRESLGPDHPAVAATLNNLAVLYGKR
GKYKAEPLCQRALEIREKVLGTHNHPDVAKQLNNLALLCQNQKYEAVERYQRALAI
```

```
FEATURES
source
1..1524
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761H229"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="amygdala"
1..1068
/gene="DKFZp761H229"
<1..1068
/gene="DKFZp761H229"
/note="similarity to kinesin light chain (Homo sapiens)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB82411.1"
/db_xref="GI:7328160"
/translation="DQNKYKAAHLLNDALSIRESLGPDHPAVAATLNNLAVLYGKR
GKYKAEPLCQRALEIREKVLGTHNHPDVAKQLNNLALLCQNQKYEAVERYQRALAI
```







QY	813	CAACCTGGTGATCCAGTACGACGCCAACGAGGTCGCTATGAGGTGGCCGTGCCACTCTGTAA	872
Db	785	CAACCTGGTGATCCAGTATGCTCGAGGGCGGTTACGAGGTGGCGGTGCCACTCTGCAA	844
QY	873	GCAGGCATAGAGGACCTGGAGCGCACATCAGGCCGTGGCCACCCCTGATGTGCCACCAT	932
Db	845	GCAGGCCCTGGAGGACCTGGAGAGACTTCGGCCACGACACCCCTGATGTGGCCACCAT	904
QY	933	GCTCAACATCCTTGCTTTGGTGTATCGTGACCAAGATAAGTAAGGAAGCTGCCACCT	992
Db	905	GCTCAACATCCTGGCCCTGGTGTACAGGGATCAGAACAAAGTATAAGACGCAGCAACCT	964
QY	993	GCTGAATGATGCCCTTAGCATCCGGGAGAGCACCTTGGGACCTGACCATCCTGTGTGGC	1052
Db	965	CCTGAACGATGCCCTGGCTATCCGTGAGAAAACCCCTGGGCGAGATCACCCCTGGGTGGC	1024
QY	1053	TGCCACACTCAACAAATTTGGCTGTGCTCTATGGCAAAAGGGCAAGTACAAGGAGGAGA	1112
Db	1025	TGCGACCTCAACAACTAGCAGTGTGTACGGTAAGCGTGGCAAGTACAAAGAGGCCGA	1084
QY	1113	GCCTCTGTGCCAGCGGCACTGGAGATTCGAGAAAAGGTCCTGGGCACGAAATCATCAGA	1172
Db	1085	GCCGCTGTGCAAAACGGCCCTGGAGATCAGGGAGAGGTTCTGGGAAAGGATCACCTGA	1144
QY	1173	TGTGGCAAAACAGCTGAACAACTGGCCCTCTTGTGCCAAAACCCAGGGCAAGTATAGGC	1232
Db	1145	TGTTGCTAAGCAACTAATAATCTGGCCTTGCTGTGCCAGACCAAGGCAAGTACGAGGA	1204
QY	1233	CGTGGAACTACTACCGAGGACACTGGCCATCTACGAGGGGCGAGCTGGGGCCGGACAA	1292
Db	1205	GGTGGAGTATTACTACGAGGGCCCTGGAGATCTACCAGCAAGCTCGGACCCGATGA	1264
QY	1293	CCCTAATGTAGCCCGGACCAAGCAAGAACAACTGGCTTCTGTACTTACCTGAACACAGGCAATA	1352
Db	1265	TCCCAACGTGGCCAAAGCAGGAATAAACCCTGGCTTCTGTATCTGAAACAAGGGAAGTT	1324
QY	1353	TGCTGAGGTGAGACACTATACAAAGAGATCCTGACCCGTGCCCATGTACAGGAGTTGG	1412
Db	1325	CAACAAGCAGAAACATTTGTACAGGAGATCTCACCCCGGCACATGAACGGGAGTTGG	1384
QY	1413	GTCTGTGGATGATGACCACAAGCCCATCTGGATGCATGCAGAGGAGCGGAGGAATGAG	1472
Db	1385	ATCTGTGGATGATGAAACAAGCCCATATGGATGCACCGGAGGAGAGAGAGGAGTGCAA	1444
QY	1473	CAAAAGCCGGCACCATGAGGGTGGACACCCCTATGCTGATGAGGCTGGTACAAGGC	1532
Db	1445	AGGAA-----AGCAGAAGGACGGGTCGTCTTTGGAGAGTATGGCGGCTGGTATAAGGC	1498
QY	1533	CTGCAAGTGAAGCAGCCCAAGTGAACAGTACTCTGAGAAACCTGGGAGCTCTGTATAG	1592
Db	1499	CTGCAAGTGGACAGTCCCAACCCCTCACAAACACCTTGAAACACCTGGGAGCACTTACCG	1558
QY	1593	GGCCAGGAAAGCTGGAGGCTGTGTAGAGCCCTGGAGGAATGTGCCCTGCGGTCCCGGAG	1652
Db	1559	CCGCCAGGGGAAGTTTGAAGCTGCGGAGACACTGGAGGAAGCTGCCCTGAGGTCACGTAA	1618
QY	1653	ACAGGGCACTGACCCCTATCAGCCAGACGAAGGTGGCAGAGCTGCTTGGGAGAGTGA	1709
Db	1619	GCAGGGTCTTGACAATGTTCAACAACACAGAGAGTGGCTGAAGTGCTAAATGACCCCTGA	1675

RESULT	4
RATKINLC	
LOCUS	RATKINLC 2386 bp mRNA
DEFINITION	Rat kinesin light chain C mRNA.
ACCESSION	M75148
VERSION	M75148.1 GI:205080
KEYWORDS	kinesin light chain.
SOURCE	Rattus norvegicus Adult Brain cDNA to mRNA.
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE	1 (bases 1 to 2386)	
AUTHORS	Cyr,J.L., Pfister,K.K., Bloom,G.S., Slaughter,C.A. and Brady,S.T.	
TITLE	Molecular genetics of kinesin light chains: Generation of isoforms by alternative splicing	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 88, 10114-10118 (1991)	
MEDLINE	92052221	
FEATURES	Location/Qualifiers	
source	1..2386	
	/organism="Rattus norvegicus"	
	/db_xref="taxon:10116"	
	/dev_stage="Adult"	
	/tissue_type="Brain"	
BASE COUNT	612 a	588 c 722 g 464 t
ORIGIN		
Query Match	32.0%;	Score 785.4; DB 12; Length 2386;
Best Local Similarity	71.7%;	Pred. No. 3.1e-176;
Matches 1074;	Conservative 0;	Mismatches 411; Indels 12; Gaps 3;
QY	216	GCTCAGCCAAAGAGAGATCCTGGGGAGCACACGCGCTGGTCAGCAAGGGCTAGAGGCCCT 275
Db	174	GCTTACGCAGGATGAATCATTTCTAAGACCAAGCAGGTGATCCAGGGCTGGAGGCCCT 233
QY	276	ACGCAGTGAACACACAGGCCCGTGTGTCGAAAGCCTGTCCCAGACCATTGAGTGTCTGCAGCA 335
Db	234	GAAGAATGAGCACAACTCCATCCTGCAGAGTCTGCTGGAGACGCTGAAGTCTTGAAGAA 293
QY	336	GGGAGGCCATGAGGAAGGGCTGTTGTCATGAGAAGGCCCGCCGACGCTTCGCCGTTCTATGGA 395
Db	294	G---GACGATGAGAGCAACCTGGTGGAGGAGAAGTCCAGCATGATCCGCAAGTCCCTTGA 350
QY	396	AAACATTGAGCTCGGGCTGAGTAGGCCCCAGGTGATGCTGGCTCTAGCCAGCCACCTGAG 455
Db	351	GATGCTGGAGCTCGCCCTGAGCGAGGCGCAGGTGATGATGGCCCTGTCTAATCACCTGAA 410
QY	456	CACAGTGGAGTCGGAGAAACAGAAAGCTCGGGCTCAGGTGGCGCGGCTATGCCAGGAGAA 515
Db	411	TGCCGTGGAGTCCGAGAAGCAGAAGCTCCGAGCTCAGGTTCCAGCTCTGTGCCAGGAGAA 470
QY	516	CCAGTGGCTCGGGGATGAGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGC 575
Db	471	CCAGTGGCTCGGGGATGAACCTGGCCAAACACACAGCAGAAGTTACAGAAGAGCGAGCAGTC 530
QY	576	TGTGGCTCAGCTGGAGGAGGAAAAGAACACCTGGAGTTCTTGGGCGAGCTGCGGCGAGTA 635
Db	531	CGTGGCTCAGCTGGAGGAGGAGAAAGACACCTGGAGTTCTATGAACCACTGAAGAATA 590
QY	636	TGATGAGGATGGACATACCTCGGAGGAGAAAGAGCGCATGCCACCAAGGATTCCTCTGA 695
Db	591	CGACGACGACATCTCTCCCTCGGAAGACAAAGACTCTGATCTTCCAAGAGACCGTTGA 650
QY	696	TGACCTCTTTCCTAATGAGGAGGAGA--GGACCCCAAGCAATGGCTTGTCCCGTGGTCA 752
Db	651	TGACCTCTTCCCGAATGATGAGGATGACCCAGGACAAGGAATCCAGCAGCAACACAGCAG 710
QY	753	AGGTGCTACAGCAGCTCAGCAGGTTGGATATGAGATCCCAGCAAGGTTGCGGACGTTGCA 812
Db	711	TGCTGCGGCGCGCCCGCCAGCAGGCGGCTACGAGATCCCTGCGCGGCTGCGCAGCGTCCA 770
QY	813	CAACCTGGTGATCCAGTACGCGACCCCAAGGTCGCTATGAGGTGGCCGCTGCCACTCTGTAA 872
Db	771	CAACCTGGTGATCCAGTATGCCCTCGCAGGGGCGTTACGAGGTGGCGGTGCCACTCTGCAA 830
QY	873	GCAGGCATAGAGGACCTGGAGCGGCACATCAGGCCGCTGGCCACCCCTGATGTGCCACCAT 932
Db	831	GCAGGCCCTGGAGGACCTGGAGAAGACTTCCGGCCACGACCAACCCCTGATGTGGCCACCAT 890
QY	933	GCTCAACATCCTTGTCTTGGTGTATCGTGACCAGAAATAAGTATAAGGAAGCTGCCACCT 992
Db	891	GCTCAACATCCTGGCCCTGGTGTACAGGGATCAGAACAAAGTATAAAGACGCAGCAACCT 950
QY	993	GCTGAATGATGCCCTTAGCATCCGGGAGAGCACCTTGGGACCTTGACCATCCTGTGTGGC 1052



Db 951 CCTGAACGATGCCCTGGCTATCCGTGAGAAACCCCTGGCCGAGATCACCCCTGCGGTGGC 1010  
QY 1053 TGCCACACTCAACAATTTGGCTGTGCTCTATGGCAAAAGGGGCAAGTACAAAGGAGGCAGA 1112  
Db 1011 TGGACCCCTCAACAACCTAGCAGTGTCTGTCAGGTAAAGCTGGCAAGTACAAAGAGGCCGA 1070  
QY 1113 GCCTCTGTGCCAGCGGGCACTGGAGATTTCGAGAAAAGGTCCTGGGCACGAATCATCCAGA 1172  
Db 1071 GCCGCTGTGCAAAACGGCCCTGGAGATCAGGAGAAGGTTCTGGGAAAGGATCACCCCTGA 1130  
QY 1173 TGTGGCAAAACAGCTGAACAACCTGGCCCTCTTGTGTGCCAAAACAGGCAAGTATGAGGC 1232  
Db 1131 TGTGTAAAGCAACTAAATAATCTGGCCCTTGTGTGCCAGAACCAGGGCAAGTACGAGGA 1190  
QY 1233 CGTGGAAACGCTACTACGAGCGAGCACTGGCCATCTACGAGGGGCGAGCTGGGGCCGGACAA 1292  
Db 1191 GGTGGAGTATTACTACGAGAGGGCCCTGGAGATCTACCAGACGAAGCTCGGACCCGATGA 1250  
QY 1293 CCCTAATGTAGCCCGACCAAGAACAACCTGGCTTCCCTGTACCTGAAACAGGGCAATA 1352  
Db 1251 TCCCAACGTGGCCCAAGCAAGAATAACCTGGCTTCCCTGTATCTGAAACAAGGGAAGTT 1310  
QY 1353 TGCTGAGGCTGAGACACTATACAAAGAGATCCTGACCCGTCGCCCATGTACAGGAGTTTGG 1412  
Db 1311 CAACAAGCAGAAACATGTACAAGGAGATTCTCACCCGCGCACATGAACGGGAGTTTGG 1370  
QY 1413 GTCTGTGGATGATGACCACAAGCCCATCTGTGATGCATGCAGAGGAGCGGGAGGAAATGAG 1472  
Db 1371 ATCTGTGGATGATGAAAACAAGCCCATATGGATGCACGCGGAGGAGAGAGGAGTGCAA 1430  
QY 1473 CAAAAGCCGGCACCATGAGGGTGGGACACCCATATGCTGAGTATGGAGGCTGGTACAAGGC 1532  
Db 1431 AGGAA-----AGCAGAAGGACGGGTGCTCTTTGGAGAGTATGGCGGCTGGTATAAGGC 1484  
QY 1533 CTGCAAAAGTGAGCAGCCCCACAGTGAACACFACCTCTGAGAAAACCTGGGAGCTCTGTATAG 1592  
Db 1485 CTGCAAAAGTGACAGTCCCACCGTCACAACCAACCTTGAAAACCTTGGGAGCACTTTACCG 1544  
QY 1593 GCGCCAGGGAAAGCTGGAGGCTGCTGAGACCCCTGGAGGAATGTGCCCTGCGGTCCCGGAG 1652  
Db 1545 CCGCCAGGGGAAGTTGAAGCTGCGGAGACACTGGAGGAAGCTGCCCTGAGGTACAGTAA 1604  
QY 1653 ACAGGGCACTGACCCCTATCAGCCAGACGAAGGTGGCAGAGCTGCTTGGGGAGAGTGA 1709  
Db 1605 GCAGGGTCTTGACAATGTTCAACAACAGAGAGTGGCTGAAGTGCTAAATGACCCCTGA 1661

RESULT 5  
RATKINLA  
LOCUS RATKINLA 2308 bp mRNA ROD 27-APR-1993  
DEFINITION Rat kinesin light chain A mRNA.  
ACCESSION M75146  
VERSION M75146.1 GI:205078  
KEYWORDS kinesin light chain.  
SOURCE Rattus norvegicus Adult Brain cDNA to mRNA.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 2308)  
AUTHORS Cyr,J.L., Pfister,K.K., Bloom,G.S., Slaughter,C.A. and Brady,S.T.  
TITLE Molecular genetics of kinesin light chains: Generation of isoforms by alternative splicing  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88, 10114-10118 (1991)  
MEDLINE 92052221  
FEATURES Location/Qualifiers  
source 1. .2308  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/dev\_stage="Adult"  
/tissue\_type="Brain"  
BASE COUNT 600 a 577 c 686 g 445 t  
ORIGIN

Query Match 32.0%; Score 783.8; DB 12; Length 2308;  
Best Local Similarity 71.7%; Pred. No. 7.6e-176;  
Matches 1073; Conservative 0; Mismatches 412; Indels 12; Gaps 3;  
QY 216 GCTCAGCCCAAGAGAGATCCTGGGAGCAGACAGCGGTGGTCAGCAAGGGGCTAGAGGCCCT 275  
Db 165 GCTTACGCGAGGATGAAATCATTTCTAAGACCAAGCAGGTGATCCAGGGGCTGGAGGCCCT 224  
QY 276 ACGCAGTGAACACACAGGCCGCTGCTGCAAAAGCCCTGTCCCAGAGCCATTGAGTGTCTGCAGCA 335  
Db 225 GAAGAATGAGCACAACTCCATCCTGCAGAGTCTGCTGGAGACGCTGAAGTGTCTTGAAGAA 284  
QY 336 GGGAGGCCATGAGGAAGGGCTGGTGCATGAGAAGGCCCGGCAGCTTCGCCGTTCTATGGA 395  
Db 285 G---GACGATGAGAGCAACCTGGTGGAGGAGAAGTCCAGCATGATCCGCAAGTCTCTTGA 341  
QY 396 AAACATTGAGCTCGGCTGAGTGAGGCCACAGGTGATGCTGGCTTAGCCAGCCACCTGAG 455  
Db 342 GATGCTGGAGCTCGCCCTGAGCGAGGCGCAGGTGATGATGGCCCTGTCTAATCACCTGAA 401  
QY 456 CACAGTGGAGTCGGAGAAAACAGAAAGCTGCGGGCTCAGGTGCGGGGCTATGCCAGGAGAA 515  
Db 402 TGCCGTGGAGTCCGAGAAGCAGAAGCTCCAGAGTCCAGGTTCCAGCTGTGCCAGGAGAA 461  
QY 516 CCAGTGGCTCGGGATGAGTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGC 575  
Db 462 CCAGTGGCTCGGGATGAACCTGGCCAACACACAGCAGAGAAGTTACAGAAGAGCAAGCAGTC 521  
QY 576 TGTGGCTCAGCTGGAGGAGGAAAAGACACCTGGAGTTCTCTGGGCGAGCTGCGGCAGTA 635  
Db 522 CGTGGCTCAGCTGGAGGAGGAGAAGAACCTGGAGTTCATGAACCAAGCTGAAGAATA 581  
QY 636 TGATGAGGATGGACATACCTCGGAGGAGAAAAGGCGGATGCCACCAAGGATTCCTCTGGA 695  
Db 582 CGACGACGACATCTCTCCCTCGGAAGACAAAGACTCTGATTTCTCCAAGAGCCGTGGA 641  
QY 696 TGACCTCTTTCCCTAATGAGGAGGAAGA---GGACCCAGCAATGGCTTGTCCCGTGGTCA 752  
Db 642 TGACCTCTTCCCGAATGATGAGGATGACCCAGGACAAGGAATCCAGCAGCAACACAGCAG 701  
QY 753 AGGTGCTACAGCAGCTCAGCAGGGTGGATPATGAGATCCCAGCAAGGTTGCGGACGTTGCA 812  
Db 702 TGCTGCGGCCGCGCCAGCAGGGCGGCTACGAGATCCCTGCGCGGCTGCGCACGCTCCA 761  
QY 813 CAACCTGGTATCCAGTACGACGCCCAAGTCGCTATGAGGTGGCGTGCCTACTCTGTAA 872  
Db 762 CAACCTGGTATCCAGTATGCCTCGCAGGGCGGTACGAGGTGGCGTGCCTACTCTGCAA 821  
QY 873 GCAGGCACTAGAGGACCTGGAGCGCACATCAGGCCCGTGGCCACCCCTGATGTCGCCACCAT 932  
Db 822 GCAGGCCCTGGAGGACCTGGAGAAGACTTCGGGCCACGACCCCTGATGTGGCCACCAT 881  
QY 933 GCTCAACATCCTTGGCTTTGGTGTATCGTGACCAGAAATAAGTATAAGGAAGCTGCCCCACCT 992  
Db 882 GCTCAACATCCTGGCCCTGGTGTACAGGGATCAGAACAAAGTATAAAGACGCAAGCAACCT 941  
QY 993 GCTGAATGATGCCCTTAGCATCCGGGAGAGCACCTTGGACCTTGGACCTGACCATCTCTGTGGC 1052  
Db 942 CTTGAACGATGCCCTGGCTATCCGTGAGAAAACCCCTGGGCCGAGATCACCTGCGGTGGC 1001  
QY 1053 TGCCACACTCAACAATTTGGTGTGCTCTATGGCAAAAGGGGCAAGTACAAGGAGGCAGA 1112  
Db 1002 TGCAGCCCTCAACAACCTAGCAGTGTGTACGGTAAGCGTGGCAAGTACAAGAGAGGCCGA 1061  
QY 1113 GCCTCTGTGCCAGCGGCACCTGGAGATTCGAGAAAAAGGTCTCTGGGCACGAATCATCCAGA 1172  
Db 1062 GCCGCTGTGCAAAACGGGCCCTGGAGATCGAGGAGAAGGTTCTGGGAAAAGGATCACCCCTGA 1121  
QY 1173 TGTGGCAAAACAGCTGAACAACCTGGCCCTCTTGTGCCAAAACCCAGGCAAGTATGAGGC 1232  
Db 1122 TGTGTAAAGCAACTAAATAATCTGGCCCTTGTGTGCCAGAACCCAGGCAAGTACGAGGA 1181



QY 1233 CGTGAACGCTACTACAGCGAGCACTGGCCATCTACGAGGGCAGCTGGGGCCGGACAA 1292  
Db 1182 GGTGGAGTATTACTACAGAGGGCCCTGGAGATCTACCAGCAAGCTCGGACCCGATGA 1241  
QY 1293 CCCTAATGTAGCCCGACCAAGAACAACCTGGCTTCCTGTTACCTGAAACAGGGCAAATA 1352  
Db 1242 TCCCAACGTGGCCCAAGACGAAGAATAACCTGGCTTCCTGTTATCTGAAACAAGGAGTT 1301  
QY 1353 TGCTGAGGCTGAGACACTATACAAAGAGATCCTGACCCCGTGCCTATGTACAGGAGTTTGG 1412  
Db 1302 CAAACAAGCAGAAACATTTGTAAGGAGATTTCTACCCCGGCACATGAACGGGAGTTTGG 1361  
QY 1413 GTCTGTGGATGATGACCAAGCCCATCTGATGCATGCAGAGGAGCGGGAGGAAATGAG 1472  
Db 1362 ATCTGTGGATGATGAAACAAGCCCATATGGATGCACGCGGAGGAGAGAGTGCAA 1421  
QY 1473 CAAAAGCCGCACCATGAGGGTGGGACACCTATGCTGAGTATGGAGCTGGTACAAGGC 1532  
Db 1422 AGGAA-----AGCAAGGACGGTCTTTTGGAGAGTATGGCGGCTGGTATAAGGC 1475  
QY 1533 CTGCAAAGTGAGCAGCCCCACAGTGAACACTACTCTGAGAAACCTGGGAGCTCTGTATAG 1592  
Db 1476 CTGCAAAGTGGACAGTCCACCGTCAACAACACCTTGAAACACCTGGGAGCACTTTACCG 1535  
QY 1593 GCGCCAGGGAAGCTGGAGGCTGCTGAGACCTGGAGGAATGTGCCCTCGCGTCCCGGAG 1652  
Db 1536 CCGCCAGGGAAGTTGAAGCTCGGAGAGACTGGAGGAGCTGCCCTGAGGTCACGTAA 1595  
QY 1653 ACAGGGCACTGACCCTATCAGCCAGCAGAGCTGGCAGAGCTGCTGGGGAGAGTGA 1709  
Db 1596 GCAGGGCTTTGACAATGTTTCAACAACAGAGAGTGGCTGAAGTGCTAAATGACCCTGA 1652

RESULT 6  
CGKINID CGKINID 2055 bp mRNA ROD 16-FEB-1998  
LOCUS Cricetulus griseus mRNA for kinesin light chain isoform D.  
DEFINITION Y14586  
ACCESSION Y14586  
VERSION Y14586.1 GI:2330584  
KEYWORDS kinesin light chain.  
SOURCE Chinese hamster.  
ORGANISM Cricetulus griseus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.  
REFERENCE 1 (bases 1 to 2055)  
AUTHORS Khodjakov,A., Lizunova,E.M., Minin,A.A., Koonce,M.P. and Gyoeva,F.K.  
TITLE A specific light chain of kinesin associates with mitochondria in cultured cells  
JOURNAL Mol. Biol. Cell 9 (2), 333-343 (1998)  
MEDLINE 98119746  
REFERENCE 2 (bases 1 to 2055)  
AUTHORS Khodjakov,A.L.  
TITLE Direct Submission  
JOURNAL Submitted (13-AUG-1997) A.L. Khodjakov, Wadsworth Center, Empire State Plaza, PO Box 509, Albany, NY 12201-0509, USA

FEATURES  
source Location/Qualifiers  
1. .2055  
/organism="Cricetulus griseus"  
/db\_xref="taxon:10029"  
/cell\_line="CHO-K1"  
/clone\_lib="Unizap (Stratagene)"  
1429. .1647  
/note="Protein sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/product="kinesin light chain isoform D"  
/protein\_id="CAA74927.1"  
/db\_xref="GI:2330585"  
/db\_xref="SPTREMBL:O35500"  
/translation="VSMSEVWNGDGTSLKRSFSKLRASIRRSSEKLVKRLKGGSA  
RDSEPRNPGASPAEPLCVENVQQPRR"

CDS

BASE COUNT 548 a 502 c 615 g 390 t  
ORIGIN  
Query Match 31.1%; Score 762.8; DB 12; Length 2055;  
Best Local Similarity 59.9%; Pred. No. 7.4e-171;  
Matches 1076; Conservative 0; Mismatches 452; Indels 12; Gaps 3;  
QY 356 TGGTGCATGAGAAGGCCCGGCAGCTTCGCCCTTCTATGGAAACATTTGAGCTCGGCTGA 415  
Db 20 TGGTGGAGGAGAAGTCCAGCATGATCCGCAAGTCTCTTGAGATGCTGGAGCTCGGCTGA 79  
QY 416 GTGAGGCCCAAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACAGTGGAGTCGGAGAAC 475  
Db 80 GTGAGGCGCAGGTGATGATGGCCCTGTCCATACCTGAATGCCGTGGAGTCGGAGAAC 139  
QY 476 AGAAGCTGCGGGCTCAGGTGCGGGGCTATGCCAGGAGAACAGTGGCTCGGGATGAGC 535  
Db 140 AGAAACTGCGGCTCAGGTCCGCCGTCTGTGCCAGGAGAACAGTGGCTCGGGATGAAC 199  
QY 536 TGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAAGAACAGGCTGTGGCTCAGCTGGAGGAG 595  
Db 200 TGGCCAACACGACGCAAGAGTTGCAGAAGAGTGAACAGTCCGTGGCTCAGCTGGAGGAG 259  
QY 596 AAAAGAACACCTGGAGTTCCTGGGGCAGCTGCGGCAGTATGATGAGGATGGACATACCT 655  
Db 260 AGAAGAAGCACCTGGAGTTCATGAACCAAGCTGAAGAATATGACGATGACATCTCTCCT 319  
QY 656 CGGAGGAGAAAGAGCGGATGCCACCAAGGATTCCTGGATGACCTCTTCTTAATGAGG 715  
Db 320 CGGAGGACAAAGACTCTGATTCTACCAAGAGGCCATTGGATGACCTCTTCCCAATGACG 379  
QY 716 AGGAAGA---GGACCCCAAGCAATGGCTTGTCCCGTGGTCAAGGTGCTACAGCAGCTCAGC 772  
Db 380 AGGATGACCCAGAACCAAGGAATCCAGCAGCAACACAGCAGCGCTGCGGCCGCCCCAGC 439  
QY 773 AGGTTGGATATGATCCCAAGGTTGCGGACGTTGCACAACCTGGTGATCCAGTACG 832  
Db 440 AGGGGGCTACGAGATCCCGGCACGCGCTGCGCAGCTCCACAACCTGGTGATCCAGTATG 499  
QY 833 CAGCCCAAGTCTGCTATGAGGTGGCGCTGCCACTCTGTAAAGCAGGCACATAGAGACCTGG 892  
Db 500 CTTACAGGGGCGTTACGAGGTGGCGGTGCCCTCTGCAAGCAGGCCCTGGAGGACCTGG 559  
QY 893 AGCGCACATCAGGCGCTGGCCACCTGATGTGCCACCCTGCTCAACATCTTGTCTTGG 952  
Db 560 AGAAGACTTCCGGCCATGACCAACCTGATGTGGCCACCCTGCTGACATCTTGGCCCTGG 619  
QY 953 TGTATCGTGACCAGAATAAGTATAAGGAAGTGCACCTGCCACCCTGCTGATGATGCCCTTAGCA 1012  
Db 620 TGTACAGGGATCAGAACAAAGTACAAAGATGCAGCCAATCTGCTGAACGACGCCCTGGCTA 679  
QY 1013 TCCGGGAGAGCACCTTGGGACCTGACCATCTGCTGTGGTGGCTGCCACACTCAACAATTGG 1072  
Db 680 TCCGTGAGAAACCTTGGCCGAGATCACCCCGGCTGGCAGCAACTCTCAACAACCTCG 739  
QY 1073 CTGTGCTCTATGGCAAAAGGGCAAGTACAAGGAGGAGCAGAGCTCTGTGCCAGCGGCAC 1132  
Db 740 CAGTACTGTATGTAACGAGGGAAGTACAAGGAGGCGGACCGGCTGTGTAACGAGCCT 799  
QY 1133 TGGAGATTCGAGAAAAGTCTCTGGGACGCAATCATCCAGATGTGGCAAAACAGCTGAACA 1192  
Db 800 TGGAGATTAGAGAAGGTTCTGGGAAAGGATCATCCGGATGTAGCAACAGCAGTTAAATA 859  
QY 1193 ACCTGGCCCTCTTGTGCCAAACACCAAGGCAAGTATGAGGCCGTGGAACGCTACTACCAGC 1252  
Db 860 ACTTGGCCTTACTGTGCCAGAACCAAGGCAAGTATGAGGAGGTGGAGTATTACTACCAGA 919  
QY 1253 GAGCACTGGCCATCTACGAGGGGCGAGCTGGGGCCCGGACAAACCTTAATGTAGCCCGGACCA 1312  
Db 920 GGGCCCTGGAGATTACCAGACAAAGCTAGGACCCCGATGATCCCAATGTGGCCCAAGACGA 979  
QY 1313 AGAACAACTGGCTTCTGTACCTGAAACAGGGGCAAAATATGCTGAGGCTGAGACACTAT 1372

```

Db 980 AGAATAACCTGGCCTCCTGTTATCTGAACAAGGAAGTTCAACAAGCAGAAACGCTGT 1039
QY 1373 ACAAGAGATCCTGACCCGTGCCATGTACAGGAGTTTGGGTCTGTGGATGATGACCACA 1432
Db 1040 ACAAGGATCCTCACCCGTGCACAGCAAGCGGAGTTTGGTTCTGTGGATGATGAAACA 1099
QY 1433 AGCCCATCTGGATGCATGCAGAGAGCGGGAGGAAATGACAAAGCCCGCACCATGAGG 1492
Db 1100 AGCCCATTTGGATGCACGCGGAAGAGAGAGAGAGTGCAAAGGAA-----AGCAGAGG 1153
QY 1493 GTGGGACACCCCTATGCTGAGTATGGAGGCTGGTACAAGGCTGCAAAAGTGAGAGCCCCA 1552
Db 1154 ACGGGACTCTTTTGGAGAGTACGGTGGCTGGTACAAAGCCTGCAAAAGTGGATAGTCCTA 1213
QY 1553 CAGTGAACACTACTCTGAGAAACCTGGGAGCTCTGTATAGCGCCACAGGAAAGCTGGAGG 1612
Db 1214 CTGTCAACAACCACTTTGAAAAACCTTGGAGCACTTTACCGACGGCAGGGGAAATTGAAG 1273
QY 1613 CTGCTGAGACCCCTGGAGGAATGTGCCCTGCGGTCCCGGAGACAGGGCACTGACCTATCA 1672
Db 1274 CTGCAGAGACATTGAAGAAGCTGCCATGAGGTCACGTAAACAGAGAGTGGCTGAAGTGC 1333
QY 1673 GCCAGACGAAGGTGGCAGAGCTGCTTGGGGAGAGTGATGTTAGAAAGGACCTCCAGGAGG 1732
Db 1334 TGAATGACCTTGAGAACATGCAGAACGGGAGGAGCCGGGAGAGTCTCCACTCGGACGTGG 1393
QY 1733 GCCCTGGACACAGTGTGAAATTTCAGGGTGGTGAAGATGCTTCTGTGGCTGTGGAGTGGT 1792
Db 1394 TCAAGTACGAGAGTG---GCCCTGACGGCGGGGAGGAGGAAGTACGATGAGCGTAGAGTGA 1450
QY 1793 CCGGGATGGCAGTGGGACCCCTGCAGAGGAGTGGCTCTCTTGGCAAGATCCGGGATGTGC 1852
Db 1451 ACGGGATGGCACTGGATCTTTAAAGCGCAGTGGCTCCTTTAGCAAACTCCGGGCTCCA 1510
QY 1853 TCCGCAGACAGTGAACCTCTTGGTGAGGAAGCTCCAGGG 1892
Db 1511 TTAGCGCAGCAGTGAGAAGCTGGTTAGGAAGCTGAAGGG 1550

RESULT 7
AF055665
LOCUS AF055665 2289 bp mRNA ROD 29-JUL-1998
DEFINITION Mus musculus kinesin light chain 1 (Klcl) mRNA, complete cds.
ACCESSION AF055665
VERSION AF055665.1 GI:3347845
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2289)
AUTHORS Rahman,A., Friedman,D.S. and Goldstein,L.S.
TITLE Two kinesin light chain genes in mice. Identification and
characterization of the encoded proteins
JOURNAL J. Biol. Chem. 273 (25), 15395-15403 (1998)
MEDLINE 98288268
REFERENCE
2 (bases 1 to 2289)
AUTHORS Rahman,A., Friedman,D.S. and Goldstein,L.S.B.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-1998) CMW-W 318, HHMI, UCSD, 9500 Gilman Dr, La
Jolla, CA 92093-0683, USA
FEATURES
Location/Qualifiers
source
1..2289
/organism="Mus musculus"
/strain="Balb c"
/db_xref="taxon:10090"
/chromosome="12"
/map="between Xyl and D12Bir10"
/dev_stage="neonatal"
/tissue_type="brain"
1..2289
/gene="Klcl"
```

```

CDS
95..1720
/gene="Klcl"
/codon_start=1
/product="kinesin light chain 1"
/protein_id="AAC27740.1"
/db_xref="GI:3347846"
/translation="MYDNMSTMVYIKEEKLENVTQDEIISKTKQVIOGLEALKNEHNS
ILQSLLETCLKKDDENLVVEEKSSMIRKSLLEMLELGLSEAQVMMAALSNNHNAVESE
KONVRAQVRRLCQENQWLDELANTQKQKSEQSVALEEKKHLEFMNQLKYYDDDD
ISPSEDKDSSKEPLDDLFPNDEDEFGQGIQHSDDAAAAAARQGYEIPARLTILNLV
IQYASQGRYEVAVPSCKQALEDEKTSQHDHPDVATMLNIALYRQDNKYKDAANLL
NDALAIREKTLGRDHPAVAATLNNLAVLYGKRKYKEAEPLCKRALEIREKVLGKDH
DVAKOLNNLALCQNOGKYEEVEYYQALGIYQTKLPDPTPNVAKTKNNLASCYLK
QGFKAETLYKEILTTRAHEAEFGSVDDENKPIIMHAEERECKGKQKDGSAFGEYGG
WYKACKVDSPTVTTTLKNLGALYRRQGFEEAETLEEAAMRSRKQGLDNVHKQVRAEV
LNDPESMEKRRSRSLNMDVVKYESGPDGEEA"
BASE COUNT 591 a 569 c 690 g 438 t 1 others
ORIGIN

Query Match 29.4%; Score 721.2; DB 12; Length 2289;
Best Local Similarity 69.5%; Pred. No. 5.6e-161;
Matches 1042; Conservative 0; Mismatches 443; Indels 15; Gaps 4;

QY 218 TCAGCCAAAGAGGAGATCCTGGGGAGCACACGGCTGTCAGCCAAAGGCTAGAGGCCCTAC 277
Db 150 TCACGCAGGATGAGATCATCTAAGACCAAGCAAGTATCCAGGGCTGGAAGCCCTGA 209
QY 278 GCAGTGAACACACAGGCCGTGCTGCAAAAGCCTGTCCCAGACCATTGAGTGTCTGCAGCAGG 337
Db 210 AGAATGAGCACAACTCCATCCTCGAGAGTTTGTGGAGACGCTGAAGTGTCTGAAGAAGG 269
QY 338 GAGGCCATGAGGAAGGGCTGGTGCATGAGAAGGCCCGSCAGCTTCGCCGTTCTATGAAA 397
Db 270 ATG---ACGAGAGCAACCTGGTGAAGAGAAATCCAGCATGATCCGCAAGTCCCTGGAGA 326
QY 398 ACATTGACCTCGGGCTGAGTGAAGCCCCAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCA 457
Db 327 TGCTGGAGCTTGGCCTGAGCGAGCGGCAGGTGATGATGGCGCTGTCCAATCACCTGAATG 386
QY 458 CAGTGGAGTCGGAGAAACAGAACTGCGGGGCTCAGGTGCGGGGGCTATGCCAGGAGAAC 517
Db 387 CTGTGGAGTCCGAGAAGCAAAACGTCGCGCTCAGGTTCGACGGCTGTGCCAGGAGAAC 446
QY 518 AGTGGCTCGGGGATGAGCTGGTGGCACCCAGCAGCGGGCTACAGCGCAGTGAACAGGCTG 577
Db 447 AGTGGCTCGGGGATGAGCTGGCCAAACACGCAGCAGAAAGTTGCAGAAGAGCGAGCAGCTCGG 506
QY 578 TGGCTCAGCTGGAGGAGGAAAGAACACCTGGAGTTCTGGGGCAGCTGCGGCAGTATG 637
Db 507 TGGCGCAGCTGGAGGAGGAGAGAAACACCTGGAGTTCATGAACCAGCTGAAGAAGTACG 566
QY 638 ATGAGGATGGACATACCTCGGAGGAGAAAGAGCGGATGACCAACGATTCCTCGATG 697
Db 567 ACGACGACATCTCCCCCTCGGAGGACAAAGACTCTGATTTCTTCCAAAGAGCCGTTGGATG 626
QY 698 ACCTCTTTCTTAATGAGGAGGAAGAGGACCCCGCAATGCGTTGTCCCGTGGTCAAGGTG 757
Db 627 ATCTCTTCCAAATGACGAG--GACGAACCAAGGACAAGGAATCCAGCACACGCGACAGTA 683
QY 758 CTACAGCAGCTCAGCAGGGTGGATATGAGATCCCAGCAAGGTTGCGGACGTTGCACAACC 817
Db 684 GTGCTCGCGCGCCAGGCAGGGCTACGAGATCCCTGCAAGGCTGCGCACGCTCCACAAC 743
QY 818 TGGTGATCCAGTACGAGCCCAAGTCGCTATGAGGTGGCGGTGCCACTCTGTAAAGCAGG 877
Db 744 TGGTGATCCAGTATGCTTACAGGGGCGTTACGAAGTGGCGGTGCCATCTTTCRAAGCAGG 803
QY 878 CACTAGAGGACCTGGAGCGCACATCAGGCCCGTGGCCACCTGATGTGCCACCATGTCTCA 937
Db 804 CCCTGGAGGATCTGGAGAAGACTTCGGGCCACGACCCCGGATGTGGCTACCATGTCTCA 863
QY 938 ACATCCTTGTGTTGGTGTATCGTGACCAAGAATAAGTATAAGGAAGCTGCCCCACCTGCTGA 997
```

Db	864	ACATCTTGGCCCTGGTGTTACAGGGATCAGAAACAAGTATAAAGATGCAGCTAACCTCTCTGA	923
QY	998	ATGATGCCCTTAGCATCCGGGAGAGCACCTTGGGACCTGACCATCCTGCTGTGGCTGCCA	1057
Db	924	ACGACGCCCTGGCTATCCCGGAGAAAACCTCGGCAGAGATCACCCCGCGTGGCAGCGA	983
QY	1058	CACITCAACAATTGGCTGTGCTCTATGGCAAAAGGGGCAAGTACAAGGAGGCAGAGCCTC	1117
Db	984	CTCTGAACAACCTAGCAGTACTGTACGGTAACGGAGGGAAGTACAAGGAGCGGAGCGCG	1043
QY	1118	TGTGCCAGCGGGCACTGGAGATTCGAGAAAAGGTCTCTGGGCACCAATCATCCAGATGTGG	1177
Db	1044	TGTGTAAACGAGCCCTGGAGATCAGGGAGAAGGTTCTGGGAAGGATCATCCTGATGTTG	1103
QY	1178	CAAAACAGCTGAACAACCTTGGCCCTCTTGTGCCAAAACCAAGGCGCAAGTATGAGGCCGTGG	1237
Db	1104	CCAAACAGTTAATAACCTTGGCCCTGCTGTGCCAGAACCAAGGGTAAGTACGAGGAGGTGG	1163
QY	1238	AACGCTACTACCAGGAGGCACTGGCCATCTACGAGGGGCAGCTGGGGCCGGA--CAACC	1294
Db	1164	AGTATTATTACCAGAGGGCCCTGGGCATCTACCAGACGAAGCTGGGGCCCGATCGTACTC	1223
QY	1295	CTAATGTAGCCCGGACCAAGAACAACCTGGCTTCCTGTTACCTGAAACAGGGCAATATG	1354
Db	1224	CCAACGTGGCCCAAGACCAAGAACAACCTGGCCCTCTGTTATCTGAAACAAGGGAAGTTCA	1283
QY	1355	CTGAGGCTGAGACACTATACAAAGAGATCCTGACCCGCTGCCCATGTACAGGAGTTTGGGT	1414
Db	1284	AGCAGGCAGAAACGCTGTACAAAGAGATTCTCACCCGCGCACACGAAGCGGAGTTTGGAT	1343
QY	1415	CTGTGGATGATGACCAACAAGCCCATCTGGATGCATGCAGAGGAGCGGGAGGAAATGAGCA	1474
Db	1344	CTGTGGACGACGAGAACAAAGCCCATCTGGATGCACGCTGAAGAGAGAGAGGA-----GT	1397
QY	1475	AAAGCCGGCACCATGAGGGTGGGACACCCCTATGCTGAGTATGGAGGCTGGTACAAAGCCT	1534
Db	1398	GCAAGGCAAGCAGAAAGGACGGGTCTGGCTTTTGGAGAGTATGGCGGCTGGTATAAAGCCT	1457
QY	1535	GCAAGTGAAGCAGCCCCACAGTGAACACTACTCTGAGAAAACCTGGGAGCTCTGTATAGGC	1594
Db	1458	GCAAGTGGACAGTCCCACCGTTCACAACCCACTTGA AAAAACCTTGGAGCACTTTACCGAC	1517
QY	1595	GCCAGGGAAGCTGGAGGCTGCTGAGACCCCTGGAGGAATGTGCCCTGGCGTCCCGGAGAC	1654
Db	1518	GGCAGGGGAAGTTTGAAGCTGCAGAGACATTGGAAGAAGCCGCCCATGAGGTCACGTAAGC	1577
QY	1655	AGGGCACTGACCCCTATCAGCCACGACGAAGGTGGCAGAGCTGCTTGGGGAGAGTGTATGGTA	1714
Db	1578	AGGGTCTTGACAAATGTTCCAAACACAGAGAGTGGCTGAAAGTGCTAAATGACCTTGAGAGCA	1637

RESULT	8
HUMKINESLC	
LOCUS	HUMKINESLC 2308 bp mRNA
DEFINITION	Homo sapiens kinesin light chain mRNA, complete cds.
ACCESSION	L04733
VERSION	L04733.1 GI:307084
KEYWORDS	kinesin light chain.
SOURCE	Homo sapiens (human).
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2308) Cabeza-Arvelaiz,Y., Shih,L.C., Hardman,N., Asselbergs,F., Bilbe,G., Schmitz,A., White,B., Siciliano,M.J. and Lachman,L.B. Cloning and genetic characterization of the human kinesin light-chain (klc) gene DNA Cell Biol. 12 (10), 881-892 (1993)
JOURNAL	DNA Cell Biol. 12 (10), 881-892 (1993)
MEDLINE	94099888
FEATURES	Location/Qualifiers 1..2308 /organism="Homo sapiens"
SOURCE	source

```

/db_xref="taxon:9606"
/cell_type="T-cell"
/tissue_type="blood"
/tissue_lib="lambda-gtil, PHA-stimulated T-cell"
277. .1986
/note="putative"
/codon_start=1
/function="membrane-bounded organelles transport"
/product="kinesin light chain"
/protein_id="AAA16576.1"
/db_xref="GI:307085"
/translation="MSTMVYIKEDKLEKLTQDEIISKTKQVIQGLEALKNEHNSILQSL
LLETCLKKDDSENLVEEKSNIIRKSLLEMLGLSEAOVMMAISNHLNAVESEKQL
RAQVRLCOENQWLRLDELANTQKQKSEQSVAQLEEEKKHLEFMNQLKKYDDDISP
EDKDTSTREPLDLDLPNDEDDPGQIQQHSAAAAAQGGYEIPARLRTLHNLVLIQ
YASQREYTVAPLCKQALEDEKTSQHDHPDVATMLNIALVYRDQNKYKDAANLLND
ALAIKREYTKGDHPAVAATLNNLAVLYKRGKYKEAREPCKRALEIREKVLGKDPDPA
AKQLNNLALCQNGKYEVEYYQRALEIYQTKLPDPPNVAKTNNLASCYLKQKQK
FKQAEITLYKEITLRAHEREFGSVDDENKPIWMHAEERECEKQKDGTSFGEYGGWK
ACKVDSPTVTTLKNLGALYRROGFEAAETLEEAAMRSRKQGLDNVHKQORVAEVLND
PENMEKRRRSRESLNVVVKYESGPDGEEVSMSEVWNGSVSGRASFCGKRQQQWPGR
RHR"

BASE COUNT      618 a      545 c      705 g      440 t
ORIGIN

Query Match      29.1%;      Score 713.4;      DB 9;      Length 2308;
Best Local Similarity 68.7%;      Pred. No. 4e-159;
Matches 1029;      Conservative 0;      Mismatches 456;      Indels 12;      Gaps 3;

QY 216 GCTCAGCCAAGAGGAGATCCTGSGGAGCACACGGCTGGTCAGCCAAAGGGCTAGAGGCCCT 275
    ||| | ||| ||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 GCTTACACAGGATGAAATTATTTCTAAGACAAAGCAAGTAATTCAGGGGCTGGAAGCTTT 377

QY 276 ACGCAGTGAACACCAGGCGGTGCTGCAAAAGCCTGTCCAGACCACTTGAGTGCTGTCAGCA 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 GAAGAATGAGCACAAATCCATTTTACAAAGTTTGCTGGAGACACTGAAGTGTTGAAGAA 437

QY 336 GGGAGGCCATGAGGAAGGGCTGGTGCCATGAGAAGGCCCGCAGCTTCGCCGTCTATGGA 395
    | | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 AGATG---ATGAAAGTAATTGGTGGAGGAGAAATCAACATGATCCGGAAGTCACTGGA 494

QY 396 AAACATTGAGTCGGGCTGAGTGAGGCCCCAGGTGATGCTGGCTCTAGCCAGCCACTGAG 455
    | | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 GATGTTGAGCTCGGCTGAGTGAGSCACAGGTTATGATGGCTTTGTCAAATCACTGAA 554

QY 456 CACAGTGGAGTCGGAGAAACAGAGCTGCGGGCTCAGGTGCGGGCGGTATGCCAGGAGAA 515
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 555 TGCTGTGGAGTCGGAGAGCAGAAACTGCGTGCGCAGGTTGCTGCTGTGCCAGGAGAA 614

QY 516 CCAGTGGCTGCGGATGAGCTGGTGCCACCCAGCAGCGGCTACAGCGCAGTGAACAGGC 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 TCAGTGGCTACGGATGAAC TGCCCAACACGACGAGCAAACTGCAGAAAGAGTGAGCAGTC 674

QY 576 TGTGGCTCAGCTGGAGGAGGAAAGAAGCACCTGGAGTTCTCTGGGCAGCTGCGGCAGTA 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 TGTGGCTCAACTGGAGGAGGAGAGAGCATCTGGAGTTTATGAATCAGCTAAAAAATA 734

QY 636 TGATGAGGATGGACATACCTCGGAGGAGAGAAAGGCGATGCCACCAAGGATTCCCTGGA 695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 TGATGACGACATTTCCCCATCCGAGGACAAAGACACTGATTTCTACCAAGAGCCCTTGA 794

QY 696 TGACCTCTTTTCATATGAGGAGGAAGA--GGACCCCAAGCAATGGCTTGTCCCGTGGTCA 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 795 TGACCTTTTCCCCAATGATGAAGACGACCCAGGCAAGGAATCCAGCAGCAGCAGCAG 854

QY 753 AGGTGCTACAGCAGCTCAGCAGGGTGATATGAGATCCCAAGAGTTTCGGACCTGCA 812
    | | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 855 TGCAGCCCGGCTGCCCAGCAGGGCGGCTACGAGATCCCGCGCGGCTGCGGACGCTCCA 914

QY 813 CAACCTGGTGATCCAGTACGACGCCCAAGGTGCTATGAGGTGGCCCTGCCACTCTGTAA 872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 915 CAACCTGGTGATCCAGTACGCCCTCGCAGGGCGGCTACGAGGTAGCTGTGCCCTCTGCA 974

```







Db 732 TCAGTATCAGCAGCAGGAGGTTATGAGATCCCTGCCCGTCTACGTACTCTACACAAC 791

QY 817 CTGGTGATCCAGTACGACGCCCAAGGTCGCTATGAGGTGGCCGTGCCACTCTGTAAAGCAG 876

Db 792 CTGGTTATACACTAGCCTCACAGAGTCGCTATGAGGTGGCTGTACCACTCTGTAAACAA 851

QY 877 GCACTAGAGACCTGGAGCGCACATCAGGCCCGTGGCCACCCCTGATGTGCGCCACCATGCTC 936

Db 852 GCTCTAGAAGATCTGGAGAAGACATCAGGTCTATGATCATCCAGATGTAGCAACAATGCTC 911

QY 937 AACATCCCTTCTGTTGGTGATCGTGACCAAGAATAAGTATAAGGAAGCTGCCACCTGCTG 996

Db 912 AACATTTCTGCTCGTATACAGGGATCAAAACAATAACAAGAGGCTGGTAACCTGCTA 971

QY 997 AATGATGCCCTTAGCATCCGGGAGAGACGCTTGGGACCTGACCATCCTGTGGCTGCC 1056

Db 972 CATGATGCTCTTGTCTATCAGGGAAAGAGGCTAGGACCAGATCATCCTGCAGTTGCAGCC 1031

QY 1057 ACACCTCAACAATTGGCTGTGCTCTATGGCAAAAGGGGCAAGTACAAGGAGGCAGACCT 1116

Db 1032 ACATTGAATAACCTAGCAGTGTGTATGTTAAGAGAGGCAAGTACAAGGAAGCAGACCT 1091

QY 1117 CTGTGCCAGGGGCACTGGAGATTCGAGAAAAGGTCCTGGSCACGAATCATCCAGATGTG 1176

Db 1092 TTATGTAAGAGAGCCCTGGAGATCAGGGAAGGTTCTTGGGAAGATCATCCTGATGTA 1151

QY 1177 GCAAAACAGCTGAACAACCTGGCCCTCTTGTGCCAAAACACAGGCAAGTATGAGGCCGTG 1236

Db 1152 GCTAAGCRACTAATAATCTGGCTCTACTCTGCCAAAATCAAGGAATAATGAAGAGTT 1211

QY 1237 GAACGCTACTACCAGGAGCAGCTGCGCATCTACGAGGGGAGCTGGGCCGCGGACACCT 1296

Db 1212 GAGTGGTACTACCAGGAGCAGCTAGAAAATCTATGAGAAGAGCTTGGTCCCGATGATCCT 1271

QY 1297 AATGTAGCCCGGACCAAGAACACCTGGCTTCTTCTGTTACCTGAAACAGGGCAAAATATGCT 1356

Db 1272 AACGTAGCCAAGACAAAGAACAACTTGTGTCAGCTTACCTCAAGCAGGGCAAGTACAAG 1331

QY 1357 GAGGTGAGACACTATACAAAGAGATCCTGACCCGTGCCATGTACAGGAGTTTGGT -- 1414

Db 1332 GCTGCAGAGACCTCTACAAAGCAGGTCCTAACCGGGCCACGAAGGAGTTTGGTCTA 1391

QY 1415 ----CTGTGATGATGACCAACAGCCCATCTGGATGTCATGAGGAGCGGGAGGAATG 1470

Db 1392 AGCGCTGATGACAAGGACAAACAACTATCTGGATGTCAGGAGGAGGAGGGAAGAAAG 1451

QY 1471 AGCAAAAGCCCGCACCATGAGGTTGGGACACCCCTATGCTGAGTATGAGGCTGGTACAAG 1530

Db 1452 GGTAAATTC-----AAGGATAACGCACCGTATGTTGATTATGTTGGTGGATGGCACA 1502

QY 1531 GCCTGCAAGTG-----AGCAGCCCCACAGTGAACACTACTCTGAGAAAAC 1575

Db 1503 GCAGCTAAAGTGGATAGTCGGTCTCGCAGTAGTCTACGGTAACAACACTACCCCTAAGA 1562

QY 1576 CTGGAGCTCTGTATAGCGCGCAGGAAAGCTGGAGGCTGTGAGACCCCTGGAGGAATGT 1635

Db 1563 CTTGGGCGCTCTACAGGAGACAAGGAAAGTATGATGCTGCAGAGATACTGGAGAGTGT 1622

QY 1636 GCCCTGCGGTCCTCGGAGACAGGCGCACTGACCCCTATACCCAGACGAAGTGGCAGAGCTG 1695

Db 1623 GCCATGAAGTCACGCAGAAATGCCCTTGGACATGGTGGCTGAGACCAAGTCCGCTGAGCTT 1682

QY 1696 CTTGGGGAG 1704

Db 1683 CTAGGCCAG 1691

RESULT 10

SUSKINLCIB SUSKINLCIB 2685 bp mRNA INV 26-OCT-1993

LOCUS Strongylocentrotus purpuratus kinesin light chain isoform 2 mRNA,

DEFINITION complete cds.

ACCESSION L10234

VERSION L10234.1 GI:161527

KEYWORDS isoform 2; kinesin light chain.

SOURCE Strongylocentrotus purpuratus (library: lambda-ZAP) unfertilized egg cDNA to mRNA.

ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 2685)

AUTHORS Wedaman,K.P., Knight,A.E., Kendrick-Jones,J. and Scholey,J.M.

TITLE Sequences of sea urchin kinesin light chain isoforms

JOURNAL J. Mol. Biol. 231 (1), 155-158 (1993)

MEDLINE 93267648

FEATURES

Location/Qualifiers

1..2685

/organism="Strongylocentrotus purpuratus"

/db\_xref="taxon:7668"

/tissue\_type="unfertilized egg"

/tissue\_lib="lambda-ZAP"

132..2165

/codon\_start=1

/product="kinesin light chain isoform 2"

/protein\_id="AAA03058.1"

/db\_xref="GI:161528"

/translation="MSGSKLSTPNNSGGGGLNLSQEQIITGTREVINGLEQLKNEHND

ILNSLYQSLKMLKDDTPGDSNLVEEKTDIIEKSLESLELGEAKVMALGHLNMVE

AEKQKLRQVRLVQENTWLRDELAATQOKLOTSEQNLAADLEVKYKHLEYMNSIKKYD

EDRTPEEASSDPLDGFPEDDDDGGQADESPQPTGSGSVAAGGYEIPARLRTL

HNLIQYASQSRVEAVPLCKOALEDEKTSQHDHPDVATMLNILALVYRDQNYKEA

GNLLHDALAIREKTLGPDHPAVALTNLNLAVLYKRGKYKEAEPLCKRALEIREKVLG

KDHPDVAKQLNNLALLCQNGKYEEVEWYQRALEIYEKKLGPDDNVAKTKNNLAAA

YLKQGYKAAAEITYKVLTRAHEREFGLSADDKDNKPIWMOAEEREKGFKFDNAPYG

DYGGWHKAAKVDSSRSPTVTTLKNLALYRQGYDAAEILECAMKSRNALDM

VRETKVRELLGDLSTDVPRSEAMAKERHRRSSGTPRHGSTESVSEKTDGSEKXRK

AKDRSRIPAGYVEIPRPPHVLVENGDKLRRSGSLKLRASVRSSTKLNLKLGKR

ESDDDDGGMKRASSMSVLPISRGNDSTPAQLSQRGVRGSHDNLSSRRQSGNF"

BASE COUNT 816 a 614 c 671 g 584 t

ORIGIN

Query Match 24.2%; Score 593.4; DB 33; Length 2685;

Best Local Similarity 64.7%; Pred. No. 1.3e-130;

Matches 977; Conservative 0; Mismatches 496; Indels 36; Gaps 5;

QY 223 CAAGAGGAGATCCTGGGAGCAGCAGCGCTGGTCAGCCCAAGGCTAGAGGCCCTACGCAGT 282

Db 192 CAGGAGCAGATCATCACAGGCAGCGGAGGTGATCAAGGCTCTGGAACAGCTCAAGAAT 251

QY 283 GAACACCAGCGCGTGTCTGCAAGCCTGTCCCAGACCATTTAGTGTCTGCAGCAGGAGGC 342

Db 252 GAACACATGACATCCTCAACAGCCTCTACCAGAGCCTCAAGATGCTCAAGAAGACACA 311

QY 343 CATGAGGAAGG---CTGGTGCATGAGAAGCCCGCCGACGCTTCGCCGTTCTATGAAAAC 399

Db 312 CCAGGTGACTCTAACCTGGTGAGGAGAAGACTGATATCATCGAGAAGTCCCTTGAGAGT 371

QY 400 ATTGAGCTCGGCTGAGTGAGGCCCGCAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACA 459

Db 372 CTGGAGCTTGGTCTGGGTGAGGTAAGGTAATGATGGCATTTGGTCACCATCTCAACATG 431

QY 460 GTGGAGTCGGAGAAACAGAGCTCGGGCTCAGGTGCGCGGCTATGCCAGGAGAACCCAG 519

Db 432 GTGGAGCCGAGAAACAGAGCTACGAGCTCAGGTTTCGAAGTTAGTCCAGGAGATACC 491

QY 520 TGGCTGCGGGATGAGCTGGCTGGCACCACCGCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTG 579

Db 492 TGGTTGAGAGATGAGCTGGCTGCCACCCCAACAGAGCTACAGACGAGTGAGCAGATCTA 551

QY 580 GCTCAGCTGGAGGAGGAAAAGAACACCTGGAGTTCTTGGGGCAGCTGCGGGCAGTATGAT 639

Db 552 GCTGATCTAGAGGTCAAAATACAGCACCTGGAGTATATGAACCTCCATCAAGAGATGAT 611

QY 640 GAGGATGGACATACCTCGGAGGAGAAAGAGGCGATGCCACCAAGGATTCCTCGGATGAC 699

Db 612 GAAGATCGCACACCAGATGAGGAGGCATCATCATCAGATCCCTTAGATCTTGGTTTCCCA 671

QY 700 CTCTTTTCCCTAATGAGGAGGAAGAGGACCCCGCAATAGGCTTGTCCTC---GTGGTCAAGGT 756

Db 672 GAAGATGATGATGGTGGTCAAGCAGACGAGAGCTATCCCCAGCCCCCAACAGGATAGTGGT 731

QY 757 GCTACAGCAGCTCAGCAGGGTGGATATGAGATCCAGCAAGGTTGCGGACGTTGCACAAC 816

Db 732 TCAGTATCAGCAGCAGCAGGAGGTTATGAGATCCCTGCCCCCTACGTACTCTACACAAC 791

QY 817 CTGGTATCCAGTACGACGCCCAAGGTCGCTATGAGGTGGCCGTGCCACTCTGTGAAGCAG 876

Db 792 CTGGTTATACAGTACGCCCTCACAGAGTCGCTATGAGGTGGCTGTACCACCTCTGTAAACAA 851

QY 877 GCACTAGAGGACCTGGAGCGCACATCAGGCCCGTGGCCACCCTGATGTGCCACCATGCTC 936

Db 852 GCTCTAGAAGATCTGGAGAAGACATCAGGTCATGATCATCCAGATGTAGCAACAATGCTC 911

QY 937 AACATCCTTGCTTTGGTGTATCGTGACCAAGATTAAGTATAAGGAAGCTGCCACCTGCTG 996

Db 912 AACATCTTGCCCTCGTATACAGGGATCAAAACAATAACAAGAGGCTGGTAACCTGCTA 971

QY 997 AATGATGCCCTTAGCATCCGGGAGAGCACCTTGGGACCTGACCATCCTCTGTGGCTGCC 1056

Db 972 CATGATGCTCTTGCTATCAGGGAAGACGCTAGGACCAGATCATCCTGCGATGCGAGCC 1031

QY 1057 ACACCTCAACAATTGGCTGTCTCTATGGCAAAAGGGCAAGTACAAGGAGGCAGAGCCT 1116

Db 1032 ACATTGAATAACCTAGCAGTGTGTATGGTAAGAGAGGCAAGTACAAGGAGCAGAACCT 1091

QY 1117 CTGTGCCAGCGGCACCTGGAGATTCGAGAAAAGGTCCTGGGCACGAATATCCAGATGTG 1176

Db 1092 TTATGTAAGAGAGCCCTGGAGATCAGGGAAGAGGTTCTTGGGAAGATCATCTGATGTA 1151

QY 1177 GCAAAACAGCTGAACAACCTGGCCCTCTTGTGCCCCAAACACGAGGCAAGTATGAGGCCGTG 1236

Db 1152 GCTAAGCAACTAAATAATCTGCTCTACTCTGCCCCAAATCAAGGAAATATGAAGAGGTT 1211

QY 1237 GAACGCTACTACCAGCGAGCACTGGCCATCTACGAGGGGCGAGCTGGGGCCGACAAACCT 1296

Db 1212 GAGTGTGCTACTACCAGCGAGCACTAGAAATCTATGAGAAGAAGCTTGGTCCCGATGATCCT 1271

QY 1297 AATGTAGCCCGGACCAAGAACACCTGGCTTCTCTGTACCTGAAACAGGGSCAAATATGCT 1356

Db 1272 AACGTAGCCCAAGACAAAGAACACCTTGTCTGCAGCTTACCTCAAGCAGGGCAAGTACAAG 1331

QY 1357 GAGGCTGAGACACTATACAAAGAGATCCTGACCCGTCGCCATGTACAGGAGTTTGGGT-- 1414

Db 1332 GCTGCAGAGACCCCTCTACAAAGCAGGTCTTAACACGAGGCCACGAAAGGAGTTTGGTCTA 1391

QY 1415 ----CTGTGGATGATGACCACAGCCCCATCTGGATGCTATGCAGAGGAGCGGGAGGAATG 1470

Db 1392 AGCGCTGATGACAAGGACAAACAACCTATCTGGATGCAGGAGGAGGAGGGAAGAAAG 1451

QY 1471 AGCAAAAGCCGGCACCATGAGGCTGGGACACCCCTATGCTGAGTATGGAGGCTGGTACAAG 1530

Db 1452 GGTAAATTC-----AAGGATAACGCACCCGTATGGTGATTATGGTGGATGGCACAAA 1502

QY 1531 GCCTGCAAGTG-----AGCAGCCCCACAGTGAACACTACTCTGAGAAAAC 1575

Db 1503 GCAGCTAAAGTGGATAGTCGGTCTCGCAGTAGTCCTACGCTAACAACACTACCCCTTAAGAAC 1562

QY 1576 CTGGGAGCTCTGTATAGGCGCCAGGGAAGCTGGAGGCTGCTGAGACCCCTGGAGGAATGT 1635

Db 1563 CTTGGGCGCTCTACAGGAGACAAAGAAAGTATGATGCTGCAGAGATACTGGAGGAGTGT 1622

QY 1636 GCCCTGCGGTCCCGGAGACAGGGGCACTGACCCCTATCAGCCAGACGAAGGTGGCAGAGCTG 1695

Db 1623 GCCATGAGTCAACGCAGAAATGCCTTGGACATGGTGGCTGAGACCAAAAGTCCCGTGAGCTT 1682

QY 1696 CTTGGGGAG 1704

Db 1683 CTAGGCCAG 1691

RESULT 11

SUSKINLCIC

LOCUS SUSKINLCIC 2712 bp mRNA INV 26-OCT-1993

DEFINITION Strongylocentrotus purpuratus kinesin light chain isoform 3 mRNA, complete cds.

ACCESSION L10235

VERSION L10235.1 GI:161529

KEYWORDS isoform 3; kinesin light chain.

SOURCE Strongylocentrotus purpuratus (library: lambda-ZAP) unfertilized egg cDNA to mRNA.

ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 2712)

AUTHORS Wedaman,K.P., Knight,A.E., Kendrick-Jones,J. and Scholey,J.M.

TITLE Sequences of sea urchin kinesin light chain isoforms

JOURNAL J. Mol. Biol. 231 (1), 155-158 (1993)

MEDLINE 93267648

FEATURES

Location/Qualifiers

1..2712

/organism="Strongylocentrotus purpuratus"

/db\_xref="taxon:7668"

/tissue\_type="unfertilized egg"

/tissue\_lib="lambda-ZAP"

132..2192

/codon\_start=1

/product="kinesin light chain isoform 3"

/protein\_id="AAA03059.1"

/db\_xref="GI:161530"

/translation="MSGSKLSTPNNSGGQGNLSQEIIITGTREVIKGLEQLKNEHND ILNSLYQLKMLKKDTPGDSNLVEEKTDIIEKSLESLELIGEAKVMMALGHLNMVE AEQKLRQVRRLVQENTWLRDELAATQKLOTSEQNLADLEVKYKHLEYMNSIKKYD EDRTPDEERSSSDPLDLGFPEDDDGGQADESYQPQTGSGSVSAAAGYEIPARLRTL HNLVIOYASQSRYEAVPLCKQALEDEKTSQHDHPDVATMLNILALVYRDQNKYKEA GNLLHDLAIREKTLGPDHPAVAATLNNLAVLYGKRGKYKEAEPLCKRALEIREKVLG KDHPDVAKQLNNLALLCQNOQKYEEVEWYYQRALEIYEKILGDDDPNVAKTKNNLAA YLQKGKYKAAETLYKQVLTTRAHEREFGLSADDDKNKPIWQAEREKEGKFKDNAPYG DYGGWHKAAKVDSRSRSPPTVTTTLKNLGLYRRQGYDAAEILEECAMKSRNALDM VRETKVRELLGQDLSTDVPRSEAMAKERHRRSSGTPRHGSTESVSYEKTDGSEESI GVAWKAKRAKDRSRSIPAGYVEIPRSPPHVLVENGDKLRRSGSLSKLRASVRRST KLLNKLKGRESDDDDGGMKRASSMSVLP SRGNDESTPAPIQLSQRGVRVSHDNLSSRQ SGNF"

BASE COUNT 823 a 617 c 682 g 590 t

ORIGIN

Query Match 24.2%; Score 593.4; DB 33; Length 2712;

Best Local Similarity 64.7%; Pred. No. 1.3e-130;

Matches 977; Conservative 0; Mismatches 496; Indels 36; Gaps 5;

QY 223 CAAGAGGAGATCCTGGGGAGGACACACGGCTGGTCAGCCAAAGGGCTAGAGGCCCTACGCAGT 282

Db 192 CAGGAGCAGATCATCACAGGCACACGGGAGGTGATCAAGGCTCTGGAACAGCTCAAGAAT 251

QY 283 GAACACCAGGCCGTGCTGCAAGCCCTGTCCCAGACCATTGAGTGTCTGCAGCAGGGAGGC 342

Db 252 GAACACAATGACATCCTCAACAGCCTCTACCAGAGCCTCAAGATGCTCAAGAGGACACA 311

QY 343 CATGAGGAAGGG---CTGGTGCATGAGAAAGCCCGCCGAGCTTCGCCGTTCTATGGAAC 399

Db 312 CCAGGTGACTCTAACCTGGTGGAGGAGAAGACTGATATCATCGAGAAGTCCCTTGAGAGT 371

QY 400 ATTGAGCTCGGGCTGAGTGAGGCCAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACA 459

Db 372 CTGGAGCTGGTCTGGGTGAGGCTAAGGTAATGATGGCATGGGTCAACATCTCAACATG 431

QY 460 GTGGAGTCGGAGAAACAGAAAGCTGCGGGCTCAGGTGCGCGGGCTATGCCAGGAAACCAG 519

Db 432 GTGGAGCGGAGAAACAGAAAGCTACGAGCTCAGGTTCCGAAGGTTAGTCCAGGAGATACC 491





Db 212 GAACGGAATCGCCGAGGTGCAGAGGACAATGAAAAAGTCGGACATGCTGCGAAAGAACAT 271

QY 393 GGAAAAACATTGAGCTCGGGCTGAGTGAGGCCAGGCTGATGCTGGCTCTAGCCAGCCACCT 452

Db 272 CGAGAACATCGAACTGGCCCTCAGCGAGGCCAGGTGATGATGCCCTGACATCCCATCT 331

QY 453 GAGCACAGTGGAGTCGGAGAAACAGAAAGCTGCGGGCTCAGGTGCGGGCTATGCCAGGA 512

Db 332 GCAGAACATCGAGGCCGAGAGGACACAAGCTTAAGACACAGGTGCGCCGCTCCACCAGGA 391

QY 513 GAACCAAGTGGCTCGGGATGAGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACA 572

Db 392 GAACGCCCTGGCTCCGCGACGAGCTGGCCAAACACTCAGCAAAAAGTTCCAGGCCATCCGACA 451

QY 573 GGCTGTGGCTCAGCTGGAGGAGGAGGAGAAAGAACCTGGAGTTCCTGGGGCAGCTGGCGCA 632

Db 452 GCTGGTGCCCAACTAGAGGAGGAGAGAACGACCTGGAGTTCATGGCGTCCGCTGAAGAA 511

QY 633 GTATGATAGGATGGACATACCTCGAGGAGAGAAAGAGGCATGCCACCAAGGATTCCTCT 692

Db 512 GTACGACGAGAAT---CAGGAGCAGGACGACGCTTCGACAAAGTCCCGCACCGATCCCGT 568

QY 693 GGATGACCTCTTTCCTAATGAGGAGGAGGAGGACCCAGCAATGGCTTGTCCTCGTGTCA 752

Db 569 GGTGAGCTGTTCCCGACGAGGAGGAGGAGGACAGACACACATGTGCGCCCACTCCGCC 628

QY 753 AGGTGCTACAGCAGCTCAGCAGGCTGGATATGATATCCAGCAAGGTTGCGGACGTTGCA 812

Db 629 CAGCCAGTTGCCCAACCAGACTTCGCGCTACGAGATTCCAGCGCTCTCCGTACTCTGCA 688

QY 813 CAACCTGGTATCCAGTACGCGACCCCAAGTCGCTATGAGTGGCGCTGCCACTCTGTAA 872

Db 689 CAACCTGGTATTCAGTATGCATCGCAGGGAAGATACGAGGTAGCCGTTCCCTCTGCAA 748

QY 873 GCAGGCATAGAGGACCTGGAGCGGCACATCAGGCCGTGGCCACCCCTGATGTCGCCACCAT 932

Db 749 GCAAGCATTTGAAGATCTCGAAAGGACAAGTGGTCAAGACCATCCCGATGTAGCAACCAT 808

QY 933 GCTCAACATCTGCTTTGGTGTATCGTGACCAAGAATAAGTATAAGGAAGCTGCCACCT 992

Db 809 GCTGAACATTTGGCTCTCGTGTACCCGCTCAGAACAAAGTATAAGGAGGCCGCCAATTT 868

QY 993 GCTGAATGATGCCCTTAGCATCCGGGAGAGCACCTTGGGACCTGACCATCTCTGTGGC 1052

Db 869 ACTCAACGACGCTCTGTGATTTCGAGGAAAGACACTGGGCGAGAACCATCCAGCCGTGGC 928

QY 1053 TGCCACACTCAACAATTTGGCTGTGCTATGGCAAAAGGGGCAAGTACAAAGGAGGACAGA 1112

Db 929 GGCCACGTTGAACAACCTTGGCCGCTCTATGGCAACGCTGGCAAGTACAAAGATGCCGA 988

QY 1113 GCCTCTGTGCCAGGGGCACTGGAGATTTCAGAAAAAGGTCTCTGGGACACGAATCATCCAGA 1172

Db 989 GCCACTGTGCAAGCGTGCTTGGAGATCCGAGAGAGGTCTCTGGGAAAGGATCATCCCGA 1048

QY 1173 TGTGGCAAAACAGCTGAACAACCTGGCCCTCTTGTGCCAAAACCCAGGGCAAGTATGAGGC 1232

Db 1049 TGTTGCCAAGCAACTTAACAATCTCGCCCTGCTCTGCCAGAATCAGGGCAAAATACGACGA 1108

QY 1233 CGTGGAACGCTACTACCAGCGGACACTGGCCATCTACGAGGGGCGACCTGGGGCCGGACAA 1292

Db 1109 GGTGAGAAGTACTACCAGCGAGCTCTCGACATCTACGAATCAAAGTTGGTCCCGATGA 1168

QY 1293 CCTAATGTAGCCCGACCAAGAACAACCTGGCTTCTGTTACCTGAAACAGGGCAATA 1352

Db 1169 TCCCAACGTGGCCAGACAAAAGATAATCTGGCCGCTGCTATCTGAAGCAAGGTAGATA 1228

QY 1353 TGCTGAGGCTGAGACACTATACAAAGAGATCTGACCCGTCGCCATGTACAGGAGTTGG 1412

Db 1229 CACCGAAGCTGAAATCCTCTATAAGCAGGCTTGTACGCGAGGCCCCACGACGTGAGTTGG 1288

QY 1413 GTCTGTGGATGATGACACAAAGCCCATCTGGATGCATGCAGAGGAGCGGAGGAAATGAG 1472

Db 1289 AGCCATCGACAGCAAAAACAAGCCCATTTGGCAGGTGGCTGAGGAGCGGTGAGGA----- 1342

QY 1473 CAAAAGCCGGCACCATGAGGTGGGACACCCCTATGCTGAGTATGGAGGCTGTACAAGGC 1532

Db 1343 GCACAAATTCGATAACAGGGGAGAACACTCCATATGCGGAGTACGGCGTTGGCATAAGGC 1402

QY 1533 CTGCAAAGTGAGCAGCCCCACAGTGAACACTACTCTGAGAAACCTGGGAGCTCTGTATAG 1592

Db 1403 CGCTAAAGTGGATTGCCCCACGGTCAACAACCACTCTAAAAAATCTGGGAGCACTTACCG 1462

QY 1593 GCGCCAGGAAAGCTGGAGGCTGCTGAGACCCCTGGAGGAATGTGCCCTGCGGTCGGGAG 1652

Db 1463 ACGTCAAGGCATGTTTGAAGCGGCCGAAACCCCTGGAGACTGTGCAATGCGGAGTAAAAA 1522

QY 1653 ACAGGCACTGACCCCTATCAGCCAGACGAAGGTGGCAGAGCTGCT 1697

Db 1523 AGAAGCCTACGATCTAGCTAAACAAACGAAGCTCTCACAATTGCT 1567

RESULT 13

DROLCKIN

LOCUS DROLCKIN 1846 bp mRNA INV 26-APR-1993

DEFINITION Drosophila melanogaster (clone pBS-13a) kinesin light chain mRNA, complete cds.

ACCESSION L11328

VERSION L11328.1 GI:157813

KEYWORDS kinesin light chain.

SOURCE Drosophila melanogaster (library: lambda-gt11 of Salvatore) adult head cDNA to mRNA.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1846)

AUTHORS Gauger,A.K. and Goldstein,L.S.B.

TITLE The Drosophila kinesin light chain: Proposed structure and interaction with kinesin heavy chain

JOURNAL Unpublished (1993)

FEATURES

Location/Qualifiers

1..1846

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/dev\_stage="adult"

/tissue\_type="head"

/tissue\_lib="lambda-gt11 of Salvatore"

1..119

/gene="Klc"

/db\_xref="FlyBase:FBgn0010235"

1..1846

/gene="Klc"

/note="Kinesin light chain"

/allele=""

/db\_xref="FlyBase:FBgn0010235"

120..1646

/gene="Klc"

/codon\_start=1

/db\_xref="FlyBase:FBgn0010235"

/product="kinesin light chain"

/protein\_id="AAA28669.1"

/db\_xref="GI:157814"

/translation="MTQMSQDEIITNTKTVLQGLEALRVEHVSIMNGIAEVQKDNEKS DMLRKNINIELGLSEAQVMALTSHLQNIKAEKHLKTQVRRHLHQENAWLRDELANT OOKFOASEQLVAOLEEKKHLEFMSVKKYDENQDDACDKSRDPPVYELFPDEENE DRHNSPTPPSOFANOTSGYEIPARLRLTNLNLVIOYASQGRYEVAVPLCKQALDLEL TSQHDHPDVATMLNLALVYRDQNKYKEAANLNDALSIRGKTLGENHPAVAATLNNL AVLKGKRGKYKDAEPLCKRALEIREKVLGKDHDPVAKQLNNLALLCQNOGKYDEVEKY VORALDIYESKLGPDPPNVAKTNNLAGCVLKGRYTEAEILYKQVLTTRAHEREFGAI DSKNKPIMQVAEREEREHFDNRENTPYGEYGMWKAAKVDSPTVTTTLKNLGLYRRQ GMFEAAETLEDCAKRSKKEAYDLAKQTKLSQLLTSNEKRRSKAIKEDLDFSEKNKP"

1649..1846

/gene="Klc"

/db\_xref="FlyBase:FBgn0010235"

3'UTR

576 a 451 c 449 g 370 t

BASE COUNT







QY	1497	GACACCTATGCTGAGTATGAGGCTGGTACAAAGGCTGCAAAAGTGAGCAGGCCACAGT	1556
Db	9425	CATCCATATGGCAGTACGGCGTGGCATAAGGCCGTAAAGTGGATTGCGCCACGGT	9484
QY	1557	GAACACTACTCTGAGAAACCTGGGAGCTCTGTATAGGCGCCAGGAAAGCTGGAGGCTGC	1616
Db	9485	CACAACCACTCTAAAAAATCTGGGAGCACTTTACCGACGTCAAGGCATGTTTGAAGCGGC	9544
QY	1617	TGAGACCTTGAGGAATGTGCCCTGCGGTCCCGGAGACAGGGCACTGACCCCTATCAGCCA	1676
Db	9545	CGAAACCTTGAAGACTGTGCAATCGGAGTAAAAAGAGAGCCTACGATCTAGCTAAACA	9604
QY	1677	GACGAAGGTGGCAGAGCTGCT	1697
Db	9605	AACGAAGCTCTCACAAATTGCT	9625
RESULT 15			
SQDSKLC 2043 bp mRNA INV 09-MAR-1994			
LOCUS Loligo pealii kinesin light chain mRNA, complete cds.			
DEFINITION L24440			
ACCESSION L24440.1 GI:403176			
VERSION kinesin light chain.			
KEYWORDS Loligo pealii.			
SOURCE Loligo pealei			
ORGANISM Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;			
REFERENCE 1 (bases 1 to 2043)			
AUTHORS Beushausen,S., Kladakis,A. and Jaffe,H.			
TITLE Kinesin light chains: identification and characterization of a			
JOURNAL family of proteins from the optic lobe of the squid Loligo pealii			
MEDLINE DNA Cell Biol. 12 (10), 901-909 (1993)			
FEATURES 94099890			
source Location/Qualifiers			
1..2043			
/organism="Loligo pealei"			
/db_xref="taxon:6621"			
/dev_stage="adult"			
/tissue_type="optic lobe"			
/tissue_lib="S. Beushausen GT-10"			
1..180			
181..1896			
/codon_start=1			
/product="kinesin light chain"			
/protein_id="AAA16578.1"			
/db_xref="GI:403177"			
/translation="MEVTQTVKSRIKKIEIGKMTALSQEEIIISNTKTVIQGLDTLK			
NEHNQIINSLTSMKTIIRKENGDTNLVEEKANILKKSVDSEIELGLGEAQVMMALANHL			
QHTAEKQKLRAQVRRLCQENAWLRDELANTQQKLMSEQKVATIEEEKHLEFMNEM			
KKYDNEAQVNEEKESQSSLDLGFDDDDGGQPEVLSPTOPSAMAQAASGCEIPA			
RLRTLNLV IQASQGRYEVAVPLCKQALEDELEKTSQHDHPDVATMLNILALVYRDQG			
KYKEAANLLDALGIREKTLGPDHPAVAATLNNLAVLYGKRGKYKDAEPLCKRALVIR			
EKVLGDHPDVAKQLNNLALCONQKGKYEVEERYIQRALEIYQKELGPDPPNVAKTGN			
NLASAYLKQGYKQAEILYKEVLTTRAHEKEFGKVDNDNKP IWMQAEEREENKAKYKDG			
APQPDYGSWLKAVKVDSPVTTLKNLGALYRRQGYEAAETLEECALRSRKSALVW			
RQTKISDVLGSDFSKGQSPKDRKRSNRDRNRDSMDSVSYEKSQGDGDEHEKSLHVG			
TSHKQ"			
1581..1689			
1690..1851			
1897..2043			
BASE COUNT 705 a 393 c 464 g 481 t			
ORIGIN			
Query Match 22.0%; Score 540.8; DB 33; Length 2043;			
Best Local Similarity 61.3%; Pred. No. 4e-118;			
Matches 910; Conservative 0; Mismatches 562; Indels 12; Gaps 2;			
QY	223	CAAGAGGAGATCCTGGGAGCACACGGCTGGTCAGCCCAAGGGCTAGAGGCCCTACGCAGT	282
Db	256	CAGGAGGAGATTATCTCCAACACAAAGACAGTTATACAGGGTCTGGATACTCTGAAGAAT	315

QY	283	GAACACGAGCCGCTGCTGCAAGCCCTGTCCCAGACCCATTGAGTGTCTGCAGCAGGAGGC	342
Db	316	GAGCATAATCAATACTCAACAGTCTCTTGACTTCTATGAAAACATTAGAAAGAGAAT	375
QY	343	CATGAGGAAGGCTGGTGCATGAGAAGGCCCGGCAGCGTTCGCCGTTCTATGGAAAACATT	402
Db	376	GGCGATACTAACTTGGTGGAGAGAAAAGCGAACATCCTCAAGAAAGTCTGTGGATTCAATT	435
QY	403	GAGCTCGGGCTGAGTGAGGCCAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACAGTG	462
Db	436	GAACCTGGTCTTGGAGAAAGCTCAGGTAATGATGGCGTCTTAATCATCTTCAGCATACT	495
QY	463	GAGTCGGAGAAAACAGAAAGCTGCGGGCTCAGGTGCGGGGCTATGCCAGGAGAACAGTGG	522
Db	496	GAGGCTGAAAAACAAAACTTCGAGCTCAAGTACGCCGACTGTGTGAGGAAAATGCCTGG	555
QY	523	CTGCGGATGAGCTGGTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGCT	582
Db	556	CTCAGAGATGAATTGGCAACACTCAACAAAACTTCAGATGAGTGAGCAAAAGGTGGCT	615
QY	583	CAGCTGGAGGAGGAAAAGAACCACTGGAGTTCTTGGGGCAGCTGCGGCAGTATGATGAG	642
Db	616	ACTATTGAAGAAGAAAACATCTAGAATTCAATTCATGAATGAATGAAGAAATATGATACA	675
QY	643	GATGGACATACCTCGGAGGAGAAAGGCGATGCCACCAAGGATTCCCTGGATGACCTC	702
Db	676	AATGAAGCCCAAGTCAATGAAGAAAAGAAAGTGAACAGTCTTCACTTGTGGGCTTT	735
QY	703	TTTCTTAATGAGGAGGAGAGAGACCCAGCAATGGCTGTCCCGTGGTCAAGGTGCTACA	762
Db	736	CCAGATGATGATGATGACGAGGCCCAACCTGAGGTTTGTACACTACCCACCCAGTGCA	795
QY	763	GCAGCTCA-----GCAGGTTGGATATGAGATCCAGCAAGGTTGCGGACGTTGCACAAC	816
Db	796	ATGGACAAGCAGCAAGTGGTGGTGTGTGAAATTCAGCAAGACTTCGTACACTACATAAC	855
QY	817	CTGCTGATCCAGTACGACGCCCAAGGTCGCTATGAGGTGGCGTGCCACTCTGTAAAGCAG	876
Db	856	CTGTATTCAATACGCCCTCTCAGGGCGTTATGAAGTGGTGTGCCACTGTGCAAGCAA	915
QY	877	GCACTAGAGGACCTGGAGCGCACATCAGGCCCTGSCCACCCTGTATGTCGCCACCATGCTC	936
Db	916	GCCCTTGAGGATTTGGAGAAAACACTAGTGGCCATGACCACCTGTATGTTGCCACTATGTTA	975
QY	937	AACATCCTTGGCTTTGGTGTATCGTGACCAAGAAATAGTATAAGGAAGCTGCCACCTGCTG	996
Db	976	AATATCTTGGCTCTCGTTTACAGAGACCAAGGCAAAATATAAGGAAGCTGCCAATCTACTG	1035
QY	997	AATGATGCCCTTAGCATCCGGGAGAGCACCTTTGGACCTGACCATCCTGCTGTGGCTGCC	1056
Db	1036	AATGATGCCCTAGGATCAGAGAGAAAACATTAGGACCTGATCACCACCTGTGTGCTGCC	1095
QY	1057	ACACTCAACAATTTGGTGTGCTCTATGGCAAAAGGGGCAAGTACAAGGAGGAGAGCCT	1116
Db	1096	ACTTTAAATAATTTGGCAGTATTATATGGGAAGCGGGGTAAGTACAAGGATGCTGAACCA	1155
QY	1117	CTGTGCCAGCGGCACCTGGAGATTCGAGAAAAGTCTCTGGCACGAATCATCCAGATGTG	1176
Db	1156	CTCTGCAAAAGAGCCCTTGTCTATCCGTGAAAAGTCTTGGAAAAGACCACCCAGATGTT	1215
QY	1177	GCAAAACAGCTGAACACCTGGCCCTCTTGTGCAAAAACAGGGCAAGTATGAGGCCGTG	1236
Db	1216	GCTAAGCAACTAAACATCTGGCTCTCTCTTTGCCAGAACCAAGGAAAATATGAAGAGGTT	1275
QY	1237	GAACGCTACTACCAGCGACCTGGCCATCTACGAGGGGAGCTGGGGCGGACAAACCT	1296
Db	1276	GAGCGTTACTACGCGTGTCTTTAGAAATTTACAAAAGGAACCTGGGACAGATGATCCA	1335
QY	1297	AATGTAGCCCGGACCAAGAACCAACCTGGCTTCTCTTACCTGAAACAGGCAAAATATGCT	1356
Db	1336	AATGTTGCAAGACCAAAAAATAATTTGGCATCAGCATATTTGAAACACAGGGAATACAAG	1395
QY	1357	GAGGCTGAGACACTATACAAAAGAGATCCTGTACCCCGTGTCCCATGTACAGGAGTTGGGTCT	1416

```

      ||||| | ||| ||||| | ||| | ||| ||||| | ||| |||||
Db 1396 CAGGCTGAAATTCCTCTATAAAGAAGTACTTACTAGAGCACATGAGAAAGAATTTGGCATA 1455
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1417 GTGGATGATGACCACCAAGCCCATCTGGATGTCATGCAGAGGAGCGGAGGAAATGAGCATA 1476
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1456 GTTGATGATGACAAACAACCAATCTGGATGCAAGCTGAAGAAAGAGAGGAA-----AAC 1509
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1477 AGCCGGCACCATGAGGGTGGGACACCCCTATGCTGAGTATGGAGGCTGGTACAAAGCCTGC 1536
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1510 AAGGCCAATACAAAGATGGAGCCCTCAACCTGACTATGGAAGTTGGCTCAAAGCTGTT 1569
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1537 AAAGTGAGCAGCCCAAGTGAACACTACTCTGAGAAACCTGGAGCTCTGTATAGGCGC 1596
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1570 AAAGTGGACAGCCCAACTGTGACTACAAACCTGAAAAAAGTTAGGGCACTTTACCGGAGA 1629
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1597 CAGGGAAGCTGGAGCTGCTGAGACCCCTGGAGGAATGTGCCCTGCCGCGAGACAG 1656
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1630 CAAGGAAATATGAAGCTGCAGAAACACTAGAGGAGTGTGCTCTAAGATCAAGAAAAAGT 1689
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1657 GGCACGTACCCCTATCAGCCAGACGAAGGTGGCAGAGCTGCTTGG 1700
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1690 GCTCTGGAAGTAGTCAGGCAGACCAAAATTTCTGATGTTCTAGG 1733
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2000, 01:35:28 ; Search time 156 Seconds  
(without alignments)  
3934.110 Million cell updates/sec

Title: US-09-036-614A-2  
Perfect score: 2453  
Sequence: 1 GTGAAGTGGTGAAGAAGGG.....AGAATCAGTNCCTTNTNNG 2453

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues  
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	262	10.7	268	1 X39587	Human secreted pro
2	229.4	9.4	811	1 X39709	Gastric cancer ass
3	205.6	8.4	717	1 X39711	Gastric cancer ass
c 4	164.2	6.7	361	1 Q59840	Human brain Expres
5	155.6	6.3	354	1 Q60050	Human brain Expres
6	143.4	5.8	356	1 V90534	EST clone BK455. N
c 7	138	5.6	763	1 X39710	Gastric cancer ass
c 8	94.8	3.9	178	1 V87408	EST clone CF89. Ne
9	77.8	3.2	799	1 V55831	Nucleotide sequenc
10	77.8	3.2	9600	1 V21683	Vector plasmid pCM
11	77.8	3.2	10596	1 Q51731	Plasmid pCisEBON f
12	77.8	3.2	10596	1 T40348	Plasmid pCisEBON f
13	77.8	3.2	10596	1 X15650	Nucleotide sequenc
c 14	72.8	3.0	32207	1 V73805	KSHV LUR DNA (nucl
c 15	72.8	3.0	137507	1 V19941	KSHV long unique c
16	65.4	2.7	795	1 V55830	FLGA insert stabl
17	52.4	2.1	2000	1 N71065	Gene encoding Plas
18	51.4	2.1	203	1 V17226	SCA2 gene CAG repe
19	51.4	2.1	203	1 V30271	Glutamine rich reg
20	50.8	2.1	234	1 Q84832	Spinocerebellar at
21	49.6	2.0	2744	1 Q98470	Mispl-containing p
22	48	2.0	195	1 Q84831	Spinocerebellar at
23	47.2	1.9	165	1 V30274	Glutamine rich reg
24	47.2	1.9	2004	1 T85356	Nephila clavipes s
25	47	1.9	2214	1 V22682	New DNA sequence i
26	47	1.9	3331	1 V22683	New DNA sequence i
c 27	45	1.8	1150	1 Q26651	PfHRP-II. New vacc
c 28	45	1.8	1150	1 Q44423	Histidine-rich pro
29	44.8	1.8	543	1 Q23092	Antigen tc-7a gene
c 30	44.6	1.8	1179	1 N70884	Plasmodium falcipa
31	44.4	1.8	1209	1 T04122	Porcine acylglucos
32	43.4	1.8	1542	1 N90025	DNA encoding antig
c 33	42.8	1.7	316	1 Q23084	Antigen tc-26h gen

34	42.6	1.7	2301	1 V20445	Human c-trk oncoge
35	42.4	1.7	4697	1 V32839	Human low density
36	42	1.7	2818	1 T10554	Sheep mammary glan
37	42	1.7	3376	1 Q75166	AF-9 cDNA. New acu
38	41.8	1.7	2338	1 Q14183	N.clavipes draglin
39	41.8	1.7	2338	1 V23249	Nephila clavipes s
40	41.8	1.7	2594	1 X07075	Human KDS2 protein
41	41.4	1.7	171	1 Q84834	Spinocerebellar at
42	41.2	1.7	154	1 Q84835	Spinocerebellar at
43	41.2	1.7	2385	1 T33590	Signal transducer
44	41.2	1.7	2462	1 V35473	Human hSK3 coding
45	41.2	1.7	3120	1 V71029	Green flourescent

ALIGNMENTS

RESULT 1	
X39587	X39587 standard; DNA; 268 BP.
ID AC	X39587;
DT	21-JUN-1999 (first entry)
DE	Human secreted protein 5', EST SEQ ID No 185.
KW	Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW	forensic; gene therapy; chromosome mapping; signal peptide;
KW	upstream regulatory sequence; cytokine activity; cell proliferation;
KW	differentiation; haematopoiesis regulation; tissue growth regulation;
KW	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW	thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS	Homo sapiens.
PN	WO9906551-A2.
PD	11-FEB-1999.
PF	31-JUL-1998; IB1235.
PR	01-AUG-1997; US-905133.
PA	(GEST ) GENSET.
PI	Duclert A, Dumas Milne Edwards J, Lacroix B;
DR	WPI; 99-153781/13.
DR	P-PSDB; Y11521.
PT	New nucleic acids encoding human secreted - proteins obtained from
PT	CDNA libraries prepared from substantia nigra, cerebellum, surrenals
PT	and fetal brain tissue
PS	Claim 1; Page 319; 434pp; English.
CC	X39440 to X39597 represent 5' expressed sequence tags (ESTs) for human
CC	secreted proteins, and encode the proteins given in Y11374 to Y11531,
CC	respectively. The proteins given represent the signal peptide and an
CC	N-terminal fragment of a secreted protein. The nucleic acid sequences
CC	can be used for producing secreted human gene products. They can also
CC	be used to develop products for diagnosis and therapy. The proteins
CC	obtained may have cytokine activity, cell proliferation/differentiation
CC	activity, haematopoiesis regulating activity, tissue growth regulating
CC	activity, reproductive hormone regulating activity, chemotactic/
CC	chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC	ligand activity, anti-inflammatory activity, tumour inhibition activity
CC	or other activities. The products can be used in forensic, gene therapy
CC	and chromosome mapping procedures. The sequences can also be used for
CC	obtaining corresponding promoter sequences. The nucleic acids encoding of
CC	the signal peptide can be used for directing extracellular secretion of
CC	a polypeptide or the insertion of a polypeptide into a membrane, or
CC	importing a polypeptide into a cell.
SQ	Sequence 268 BP; 59 A; 63 C; 104 G; 40 T;

Query Match 10.7%; Score 262; DB 1; Length 268;  
Best Local Similarity 98.9%; Pred. No. 5.1e-57;  
Matches 262; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY	399	CATTGAGCTCGGGCTGAGTGAGGCCAGGTGATGCTGGCTCTAGCCACCTGAGCAC	458
Db	1	CATTGAGCTCGGGCTGAGTGAGGCCAGGTGATGCTGGCTCTAGCCASNCACCTGAGCAC	60
QY	459	AGTGGAGTCGGAGAAAACAGAGCTGCGGGCTCAGGTGCGGGCTATGCCAGGAGAACCA	518
Db	61	AGTGGAGTCGGAGAAAACAGAGCTGCGGGCTCAGGTGCGGGCTATGCCAGGAGAACCA	120







CC locations of disease-associated genes, for identification of tissue  
CC type, and for prepn. of antisense sequences, probes and constructs.  
CC EST02032 has a "excellent" coding probability as evaluated using the  
CC coding-region prediction program CRM. See also Q59041-Q61440.  
SQ Sequence 354 BP; 93 A; 104 C; 99 G; 54 T;

Query Match 6.3%; Score 155.6; DB 1; Length 354;  
Best Local Similarity 68.4%; Pred. No. 3.4e-30;  
Matches 242; Conservative 0; Mismatches 108; Indels 4; Gaps 2;

Qy 586 CTGGAGGAGGAAAAGAACCTGGAGTTCTCTGGGGAGCTGCGGACAGTATGATGAGGAT 645  
Db 1 CTGGAGGAGGAGAAGAACCTGGAGTTTATGAATCAGCTAAAAAATATGATGACGAC 60

Qy 646 GGACATACCTCGGAGGAGAAAGGCGATGCCACCAAGGATTCCTGGATGACCTCTTT 705  
Db 61 ATTTCCCATCCGAGGAGACAAAGACACTGATTCTACCAAGAGCCTCTGGATGACCTTTT 120

Qy 706 CCTAATGAGGAGGAAGA--GGACCCCGAGCAATGGCTTGTCCCGTGGTCAAGTGCTACA 762  
Db 121 CCCAATGATGAAGACGACCCCGGCAAGGAATCCAGGACGACACAGCAGTGCAGCCGCG 180

Qy 763 GCAGCTCAGCAGGGTGGATATGATATCCAGCAAGTTGCGGACGTTGCACAACTGGTG 822  
Db 181 GCTNCCAGCAGGGCGGTACGAGATCCCCGGCGGTGCGGACGCTCCACAACTGGTG 240

Qy 823 ATCCAGTACGAGCCCAAGGTCGCTATGAGGTGGCGGTGCCACTCTGTAAAGAGGCACTA 882  
Db 241 ATCCAGTACGNCTCGCAGGGCGCTACGAGGTAGCTGTGCCCTNTNCAAGCAGGCCCTG 300

Qy 883 GAGGACCTGGAGCGCACATCAGGCCGTG-GCCACCCCTGATGTGCCACCATGCT 935  
Db 301 GAGGACCTGGAGAAGACTTCAGGACACGACCCACCCGACGTGGCCACCATGCT 354

RESULT 6  
V90534  
ID V90534 standard; cDNA; 356 BP.  
AC V90534;  
DT 15-FEB-1999 (first entry)  
DE EST clone BK455.  
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
KW gene therapy; ss.  
OS Homo sapiens.  
PN WO9845436-A2.  
PD 15-OCT-1998.  
PF 10-APR-1998; U06955.  
PR 10-APR-1997; US-838821.  
PA (GEM) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M;  
PI WPI; 99-070077/06.  
DR New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
PS Claim 1; Page 571; 618pp; English.  
CC The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
SQ Sequence 356 BP; 88 A; 103 C; 124 G; 41 T;

Query Match 5.8%; Score 143.4; DB 1; Length 356;  
Best Local Similarity 68.0%; Pred. No. 4.1e-27;  
Matches 217; Conservative 0; Mismatches 96; Indels 6; Gaps 1;

Qy 1426 GACCACAAGCCCATCTGGATGCTATGTCAGAGGAGGAGGAGAAATGAGCAAAAGCCGGCAC 1485  
Db 3 GACAACAAGCCCATCTGGATGTCACGACAGGAGGAGGAGGAGAAAGCAAGGATAAGCGCCG- 61

Qy 1486 CATGAGGGTGGGACACCCCTATGCTGAGTATGGAGGCTGGTACAAGGCCTGCAAAAGTGAGC 1545  
Db 62 -----GGACAGCGCCCTATGGGGAATACGGCAGCTGTACAAGGCCTGTAAAGTAGAC 116

Qy 1546 AGCCCAACAGTGAACACTACTCTGAGAAACCTGGGAGCTCTGTATAGCGCCAGGAAAG 1605  
Db 117 AGCCCAACAGTGAACACTACTCTGCGCAGCTTGGGGGCCCTATACCGGCGCAGGGCAAG 176

Qy 1606 CTGGAGGCTGCTGAGACCCCTGGAGGAATGTGCCCTCGGTCGGGAGACAGGCGCACTGAC 1665  
Db 177 CTGGAAGCCCGCACACACTAGAGGACTGTGCCAGCGCTAACCGCAAGCAGGGTTTGGAC 236

Qy 1666 CCTATCAGCCAGACGAAGTGGCAGAGCTGCTTGGGAGAGTGATGGTAGAGGACCTCC 1725  
Db 237 CCGGCAAGCCAGACCAAGTGGTAGAACTGCTGAAAGATGGCAGTGGCAGCGCGGGGAGAC 296

Qy 1726 CAGGAGGGCCCTGGAGACA 1744  
Db 297 CGCCGACGAGCCGAGACA 315

RESULT 7  
X39710/c  
ID X39710 standard; DNA; 763 BP.  
AC X39710;  
DT 02-JUL-1999 (first entry)  
DE Gastric cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
OS Homo sapiens.  
PN WO9904265-A2.  
PD 28-JAN-1999.  
PF 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 10-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,  
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O;  
DR WPI; 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67; Page 519; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 763 BP; 191 A; 205 C; 158 G; 204 T;





D <sub>b</sub>	183	CAGGACGAGGAGGGGCAGGACGAGGAGGGGCAGGAGGGGCAGGACGAGGAGGAG	242
Q <sub>y</sub>	355	CTGGTGCAAGAAAGCCCGCAGCTTCGCCGTCTATGGAACAATTGAGCTCGGCTG	414
D <sub>b</sub>	243	GGCAGGAGGGCAGGAGGGGCAGGACGAGGAGGGGCAGGAGGAGGAGGGGCAG	302
Q <sub>y</sub>	415	AGTGAGGCCAGGTGATGCTGGCTCTAGCCAGCCACCCTGACACAGTGGAGTCGAGAAA	474
D <sub>b</sub>	303	GAGGGCAGGACGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGACGAGGAGGGG	362
Q <sub>y</sub>	475	CAGAAGCTCGGGCTCAGGTGCGGGCGCTATGCCAGGAGAACCATGGC--TGCGGGAT	531
D <sub>b</sub>	363	CAGGACGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGAG	422
Q <sub>y</sub>	532	GAGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGCTCAGCTGGAG	591
D <sub>b</sub>	423	GAGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGAGGAGGGGCAGGAG	482
Q <sub>y</sub>	592	GAGGAAAGAACACCTGGAGTTCTTGGGGCAGCTGCGGCAGTATGATGAGGATGGACAT	651
D <sub>b</sub>	483	CAGGAGGGGCAGGACGAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGGGCAG	542
Q <sub>y</sub>	652	ACCTCGGAGGAGAAAGAGCGATGCCACCAAGGATTCCCTTGGATGACCTCTTTCCTAAT	711
D <sub>b</sub>	543	GAGCAGGAGGGGCAGGAGGGGCAGGACGAGGAGGGGCAGGAGGGGCAGGAGGAGGAG	602
Q <sub>y</sub>	712	GAGGAGGAAGAGG	724
D <sub>b</sub>	603	GAGGGGCAGGAGG	615

RESULT	10
V21683	
ID	V21683 standard; DNA; 9600 BP.
AC	V21683;
DT	17-AUG-1998 (first entry)
DE	Vector plasmid pCMVKmITR-EPI.
KW	Polynucleotide delivery; plasmid pCMVKmITR-EPI; vector;
KW	gene therapy; vaccine; polycationic agent; ss.
OS	Chimeric - Epstein-Barr virus.
OS	Chimeric - Adeno-associated virus.
OS	Chimeric - Cytomegalovirus.
OS	Chimeric - Bos taurus.
FH	key
FT	Location/Qualifiers
FT	14. .2594
FT	/*tag= a
FT	/product= "EBV nuclear antigen A"
FT	2623. .4559
FT	/*tag= b
FT	/note= "EBV origin of replication"
FT	4928. .5104
FT	/*tag= c
FT	/rpt_type= INVERTED
FT	/note= "AAV inverted terminal repeat"
FT	7189. .7355
FT	/*tag= d
FT	/rpt_type= INVERTED
FT	/note= "AAV inverted terminal repeat"
FT	5112. .6734
FT	/*tag= e
FT	/note= "CMV immediate-early enhancer/promoter"
FT	6818. .7050
FT	/*tag= f
FT	/note= "bovine growth hormone polyA sequence"
PN	WO9806437-A2.
PD	19-FEB-1998.
PF	13-AUG-1997; U14465.
PR	13-AUG-1996; US-023867.
PA	(CHIR ) CHIRON CORP.
PI	Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA,
PI	Murphy JE, Tetsuo U, Zukermann R;
DR	WPI: 98-159296/14.

PT Polycationic agents based on alpha-amino acids, able to complex  
 PT with nucleic acid - to facilitate its entry into cell, condense it  
 PT and protect it against serum degradation, particularly for use in  
 PT gene therapy  
 PS Disclosure; Page 77-80; 100pp; English.  
 CC This polynucleotide comprises the DNA sequence of vector plasmid  
 CC pCMVkmITR-EPI, which contains an Epstein-Barr virus (EBV) origin  
 CC of replication from plasmid pCEP4, a coding region for EBV nuclear  
 CC antigen A from pCEP4, a pair of inverted terminal repeats from  
 CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a  
 CC bovine growth hormone polyA sequence, and a kanamycin resistance  
 CC selectable marker. Polynucleotides encoding polypeptides, such as  
 CC erythropoietin or leptin, and ribozymes and antisense  
 CC polynucleotides can be inserted into the vector. The vector is  
 CC preferred for use in novel compositions and methods for improved  
 CC polynucleotide delivery into cells. In these methods, polycationic  
 CC agents are used to increase the frequency of uptake of a  
 CC nucleic acid (see also V21684-86) into a cell. The polycationic  
 CC agent can condense with the nucleic acid and inhibit serum and/or  
 CC nuclease degradation of the nucleic acid. The nucleic acid can be  
 CC a vector, may express a therapeutic protein or a vaccinating viral  
 CC or cancer antigen, or is itself therapeutic (antisense or  
 CC ribozyme). The methods and compositions can be used in the gene  
 CC therapy of many diseases.  
 SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T;

Query Match 3.2%; Score 77.8; DB 1; Length 9600;  
 Best Local Similarity 47.6%; Pred. No. 4.4e-10;  
 Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps

RESULT 11









```
QY 592 GAGGAAAAGAACACCTGGAGTTCCTGGGGCAGCTGGGGCAGTATGATGAGGATGGACAT 651
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2688 CAGGAGGGGCAGGACGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2747

QY 652 ACCTCGGAGGAGAAAGACCGCATGCCACCAAGGATTCCTGGATGACCTCTTTCTTAAT 711
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAG 2807

QY 712 GAGGAGGAAGAGG 724
      ||||| |||||
Db 2808 GAGGGGCAGGAGG 2820

RESULT 14
V73805/c
ID V73805 standard; DNA; 32207 BP.
AC V73805;
DE 25-FEB-1999 (first entry)
DE KSHV LUR DNA (nucleotides 105,301-137,507).
KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bel-2;
KW dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;
KW glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2;
KW v-adh; G-protein coupled receptor; FGARAT; ds.
OS Kaposi's sarcoma-associated herpesvirus.
PN US5849564-A.
PD 15-DEC-1998.
PF 29-NOV-1996; 770379.
PR 29-NOV-1996; US-770379.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
DR WPI; 99-069741/06.
PT Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT di:hydro:folate reductase and is useful for treatment, prophylaxis
PT or diagnosis of Kaposi's sarcoma
PS Disclosure; Column 155-182; 109pp; English.
CC This sequence is a fragment of the Kaposi's sarcoma-associated
CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67
CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,
CC ORF69, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D,
CC ORF73 which encodes immediate early protein (IEP), K14 which encodes
CC OX-2 (v-adh), ORF74 which encodes G-protein coupled receptor, ORF75
CC which encodes tegument protein/FGARAT, K15. KSHV is a new human
CC Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the
CC most common form of neoplasm occurring in persons with acquired immune
CC deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,
CC prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma
CC and for detecting expression of a DNA virus associated with Kaposi's
CC sarcoma in a cell.
SQ Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T;
```

Query Match 3.0%; Score 72.8; DB 1; Length 32207;  
Best Local Similarity 44.9%; Pred. No. 1.2e-08;  
Matches 275; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

```
QY 282 TGAACACACGAGCGCTGCTGCAAGCCCTGTCCAGACCATTTAGTGTCTGCAGCAGGAGG 341
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20023 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGAGGA 19964

QY 342 CCATGAGGAAGGGCTGGTGCATGAGAAGGCCCGGCAGCTTCGCCGTTCTATGGAAAACAT 401
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19963 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGAGGA 19904

QY 402 TGAGCTCGGGCTAGTGAGGCCCGCCAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACAGT 461
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19903 TGAGCAGGAGCAGCAGGATGAGCAGGAGGAGCAGCAGGATGAGCAGCAGGATGAGCA 19844

QY 462 GGAGTCGGAGAAACAGAAGCTGCGGGCTCAGTGCAGCGGCTATGCCAGGAGAACCAAGTG 521
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19843 GCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCA 19784
```

```
QY 522 GCTCGGGGATGAGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGC 581
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19783 GGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGGATGAACAGGAGCAGCAGGAGGA 19724

QY 582 TCAGCTGGAGGAGGAAAAGAACACACCTGGAGTTCCTGGGGCAGCTGCGGCAGTATGATGA 641
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19723 GCAGAGCAGCAGGAGGAGGAGCAGGAGCAGGAGTTCAGAGGAGCAGGAGCAGGAGTTCAGGA 19664

QY 642 GGATGGACATACCTCGGAGGAGAAAGAGCCGATGCCACCAAGGATTCCTCGGATGACCT 701
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19663 TCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGA 19604

QY 702 CTTTCCTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 761
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19603 GGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTT 19544

QY 762 AGCAGCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 821
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19543 AGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 19484

QY 822 GATCCAGTACGAGCCCAAGGTCGCTATGAGGTGCGCGTCCACTCTGTAAAGCAGGCACT 881
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19483 GTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 19424

QY 882 AGAGGAGCCTGGA 893
      ||||| |||||
Db 19423 GGAAGAGCAGGA 19412

RESULT 15
V19941/c
ID V19941 standard; DNA; 137507 BP.
AC V19941;
DE 03-AUG-1998 (first entry)
DE KSHV long unique coding region and terminal repeat.
KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW complement-binding protein; glycoprotein; capsid protein IV; infection;
KW immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
KW lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW HIV immune status; anti-inflammatory agent; therapy; ds.
OS Kaposi's sarcoma-associated herpes virus.
FH Key Location/Qualifiers
FT CDS 1142..2794
FT /tag= a
FT /product= complement-binding protein
FT CDS 8699..11236
FT /tag= b
FT /product= glycoprotein B
FT CDS complement (17261..17875)
FT /tag= c
FT /product= interleukin 6
FT CDS complement (21548..21832)
FT /tag= d
FT /product= macrophage inflammatory protein II
FT CDS complement (27137..27424)
FT /tag= e
FT /product= interferon regulatory factor 1
FT CDS 28661..29741
FT /tag= f
FT /product= protein T1.1
FT CDS complement (58976..60175)
FT /tag= g
FT /product= glycoprotein M
FT CDS complement (69412..69915)
FT /tag= h
FT /product= glycoprotein L
FT CDS complement (88410..88910)
FT /tag= i
FT /product= interferon regulatory factor 2
FT CDS 89600..90541
FT /tag= j
FT /product= interferon regulatory factor 3
```

```
FT CDS          90173..90643
FT FT           /*tag= k
FT /product= glycoprotein X
FT complement'(93636..94127)
FT FT           /*tag= l
FT /product= interferon regulatory factor 4
FT complement (111931..112443)
FT FT           /*tag= m
FT /product= capsid protein IV
FT complement (123808..127296)
FT FT           /*tag= n
FT /product= immediate early protein
PN W09804576-A1.
PD 05-FEB-1998.
PF 22-JUL-1997; U13346.
PR 29-NOV-1996; US-757669.
PR 25-JUL-1996; US-686243.
PR 25-JUL-1996; US-686349.
PR 25-JUL-1996; US-686350.
PR 25-JUL-1996; US-687253.
PR 25-JUL-1996; US-688814.
PR 05-SEP-1996; US-708678.
PR 10-OCT-1996; US-728323.
PR 13-NOV-1996; US-747887.
PR 13-NOV-1996; US-748640.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
PT WPI; 98-130615/12.
DR New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT proteins - useful for, e.g. detecting levels of HHV8 in, and
PT preparation of vaccines for treatment of, HIV patients
PT Example 2; Page 135-203; 230pp; English.
PS This sequence represents the long unique region and terminal repeat of
CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC invention which encode KSHV polypeptides selected from: (a) viral
CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC by it, and antibodies (Ab) specific for the proteins are useful for
CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC fluids or tissue samples. HHV8 infections can be treated with antisense
CC or triplex forming molecules or agents that bind specifically to the
CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC while the protein can be used in protective vaccines. Ab may also be used
CC to differentiate between lymphomas, and HHV8 may be implicated in many
CC other lymphoproliferative diseases such as lymphomas, leukaemia,
CC splenomegaly and mycosis fungoides. Cells and animals containing the
CC nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC can be inhibited with methotrexate. These can also be used to determine
CC the immune status of a patient infected with HIV. HHV8 derived protein
CC viral MIP III may be used as an anti-inflammatory agent for,
CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC 81 open reading frames.
SQ Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T;
```

Query Match 3.0%; Score 72.8; DB 1; Length 137507;  
Best Local Similarity 44.9%; Pred. No. 2.le-08;  
Matches 275; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

```
QY 282 TGAACACACGCGCTGCTGCAAGCCCTGCCAGACCATTGAGTGCTGCAGCAGGGAGG 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125323 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGA 125264

QY 342 CCATGAGGAAGGCTGGTGATGAGAGGCCCGGCAGCTTCGCCGTTCTATGGAAACAT 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125263 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGA 125204

QY 402 TGAGCTCGGGCTGAGTGAGGCCCGCAGGTGATGGCTCTAGCCAGCCACCTGAGCACAGT 461
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 125203 TGAGCAGGACGACGAGGATGAGCAGGAGCAGCAGGATGACAGGAGCAGCAGGATGAGCA 125144
QY 462 GGAGTCGGAGAAACAGAAAGCTCGGGGCTCAGGTGCGGGGCTATGCCAGGAGAACCACTG 521
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125143 GCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCA 125084
QY 522 GCTGCGGGATGAGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGAGGTGTGGC 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125083 GGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAACAGGAGCAGCAGGAGGA 125024
QY 582 TCAGCTGGAGGAGAAAGACCTCGGAGTTCCCTGGGGAGCTCGGGCAGTATGATGA 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125023 GCAGGAGCAGCAGGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGA 124964
QY 642 GGATGGACATACCTCGGAGGAGAAAGGCGATGCCACCAAGGATTCCCTGGATGACCT 701
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124963 TCAGGAGCAGGAGTTAGAGGAGCAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGA 124904
QY 702 CTTTCCTAATGAGGAGGAAGAGAGGCCCAAGCAATGGCTTGTCCCGTGGTCAAGGTGCTAC 761
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124903 GGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAGTT 124844
QY 762 AGCAGCTCAGCAGGGTGGATATGATATCCAGCAAGTTGCGGACGTTGCACAACTGGT 821
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124843 AGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCAGGAGCA 124784
QY 822 GATCCAGTACGACGCCCAAGGTGCTATGAGGTGGCCGTCGCACTCTGTAAGCAGGCACT 881
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124783 GTTAGAGGAGCAGGAGGTGGAAGAGCAAGAGCAGGAGGTGGAAGAGCAAGAGCAGGAGCA 124724
QY 882 AGAGGACCTGGA 893
    ||| ||| |||
Db 124723 GGAAGAGCAGGA 124712
```

Search completed: August 15, 2000, 10:46:27  
Job time: 33059 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:	August 15, 2000, 00:40:42 ; Search time 52.1 Seconds (without alignments) 6476.477 Million cell updates/sec
---------	---

Title:	US-09-036-614A-2
Perfect score:	2453
Sequence:	1 GTGAAGTGGTGAAGAAGGG.....AGAATCAGTNCCTTNTNTNNG 2453

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters:	486160
--	--------

```
Minimum DB seq length: 0
Maximum DB seq length: 1000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 summaries

```
Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %		Length	DB	ID	Description
		Match	%				
C	1	77.8	3.2	5452	4	US-09-130-114-1	Sequence 1, Appl
	2	77.8	3.2	10596	1	US-07-884-811-15	Sequence 15, Appl
	3	77.8	3.2	10596	1	US-07-885-971-15	Sequence 15, Appl
	4	77.8	3.2	10596	1	US-08-087-783A-15	Sequence 15, Appl
	5	77.8	3.2	10596	1	US-08-194-088B-15	Sequence 15, Appl
	6	77.8	3.2	10596	3	US-08-194-087-15	Sequence 15, Appl
	7	77.8	3.2	10596	6	PCT-US93-04648-15	Sequence 15, Appl
8	72.8	3.0	3489	4	US-08-728-323A-1	Sequence 1, Appl	
C	9	72.8	3.0	32207	3	US-08-770-379-20	Sequence 20, Appl
C	10	59.6	2.4	7218	1	US-08-232-463-14	Sequence 14, Appl
	11	55.6	2.3	9551	1	US-08-056-200-93	Sequence 93, Appl
	12	55.6	2.3	9551	4	US-08-800-644-93	Sequence 93, Appl
	13	50.8	2.1	234	2	US-08-469-802B-3	Sequence 3, Appl
	14	50.8	2.1	234	3	US-08-267-803B-3	Sequence 3, Appl
	15	48	2.0	195	2	US-08-469-802B-2	Sequence 2, Appl
	16	48	2.0	195	3	US-08-267-803B-2	Sequence 2, Appl
	17	47	1.9	2214	5	US-08-864-038A-1	Sequence 1, Appl
	18	47	1.9	3331	5	US-08-864-038A-2	Sequence 2, Appl
	19	47	1.9	3331	5	US-08-864-038A-4	Sequence 2, Appl
20	46.4	1.9	2793	2	US-08-209-747-1	Sequence 4, Appl	
21	46.4	1.9	2793	2	US-08-458-298-1	Sequence 1, Appl	
22	45.8	1.9	7218	1	US-08-232-463-14	Sequence 14, Appl	
C	23	45	1.8	1150	1	US-08-161-406-1	Sequence 1, Appl
	24	44.8	1.8	543	7	5273901-6	Patent No. 5273901
	25	44.4	1.8	1209	2	US-08-553-703A-5	Sequence 5, Appl
26	44.4	1.8	1209	2	US-08-553-703A-9	Sequence 9, Appl	

## ALIGNMENTS

RESULT 1  
 US-09-130-114-1/C  
 ; sequence 1, Application US/09130114  
 ; Patent No. 5976807  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Horlick, Robert A.  
 ; APPLICANT: Damaj, Bassam B.  
 ; APPLICANT: Robbins, Alan K.  
 ; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
 ; TITLE OF INVENTION: From Multiple Transfected Episomes  
 ; FILE REFERENCE: 0867/1D903US1  
 ; CURRENT APPLICATION NUMBER: US/09/130,114  
 ; CURRENT FILING DATE: 1998-08-06  
 ; NUMBER OF SEQ ID NOS: 36  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 5452  
 ; TYPE: DNA  
 ; ORGANISM: VEBNA  
 US-09-130-114-1

		Query Match	3.2%; Score 77.8; DB 4; Length 5452;
		Best Local Similarity	47.6%; Pred. No. 1.4e-10;
		Matches 263; Conservative	0; Mismatches 287; Indels 3; Gaps 1;
<hr/>			
QY	175	GCCCTGGTGTGCGGCAGCGGGATGAGCCTGCAGGCCACCGGCTCAGCCAAGAGGAGTC	234
Dd	2117	GGCAGGAGCACAGGAGGCGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAG	2058
<hr/>			
QY	235	CTGGGGAGCACACGGCTGTCTCAGCCAAGGGCTAGAGGCCCTACGAGTGAAACACCAGGCC	294
Dd	2057	GAGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGG	1998
<hr/>			
QY	295	GTGCTGCAAGAAGCCCTGTCCAGACCATTTAGTGTCTGCAGCAGSAGGCCCATGAGGAAGGG	354
Dd	1997	CAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGAGGGSCAGGAGSGGCAGGAGGAGG	1938
<hr/>			
QY	355	CTGGTGCATGAGAAGGCCCGGCAGCTTCGCCGTTCTATGGAACAATTGAGCTCGGGCTG	414
Dd	1937	GGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGAGGGGCAG	1878
<hr/>			
QY	415	AGTAGGCCCCAGGTGATGCTGGCTCTAGCCAGCCCACCTGAGCACAGTGGAGTCGGAGAAA	474
Dd	1877	GAGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGG	1818
<hr/>			
QY	475	CAGAAGCTCGGGCTCAGTTCGGCGGGCTATGCCAGGAGAACAGTGGC--TGCGGGAT	531
Dd	1817	CAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGG	1758





```

; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 755D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2614
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-194-088B-15

Query Match          3.2%; Score 77.8; DB 1; Length 10596;
Best Local Similarity 47.6%; Pred. No. 1.8e-10;
Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

QY 175 GGCCTGGTGTGGGGCAGCGGGATGAGCCTGCAGGCCACCGCTCAGCCCAAGAGGAGATC 234
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2268 GGGCAGGACGAGGAGGAGGGGCAGGAGGAGGAGGGGGCAGGAGGGGCAGGAGGGGCAG 2327

QY 235 CTGGGGAGCACACGGCTGGTCAGCCAAAGGGCTAGAGGCCCTACGCAGTGAACACAGGCC 294
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2328 GAGGGGCAGGACGAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGG 2387

QY 295 GTGCTGCAAGCCTGTCCCAGACCATTGAGTGTCTGCAGCAGGGAGGCCATGAGGAAGGG 354
    | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2388 CAGGAGCAGGAGGAGGGGCAGGACGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAG 2447

QY 355 CTGGTGCATGAGAAGGCCCGGAGCTTCGCCGTTCTATGGAACATTGAGTCGGGCTG 414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2448 GGGCAGGAGGGGCAGGAGGGGCAGGACGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2507

QY 415 AGTGAGGCCACAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACAGTGGAGTCGGAGAAA 474
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2508 GAGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGG 2567

QY 475 CAGACGTGCGGGCTCAGGTGCGGGCGGCTATGCCAGGAGAACCAGTGGC---TCCGGGAT 531
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2568 CAGGACGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGAG 2627

QY 532 GAGCTGGCTGGCACCCAGCAGCGGGCTACAGCGGCAGTGAACAGGCTGTGGCTCAGCTGGAG 591
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG 2687

QY 592 GAGGAAAAAGACACCTGGAGTTCCTGGGGCAGCTCGGGCAGTATGATGAGGATGGACAT 651
```

```

Db 2688 CAGGAGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2747
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 652 ACCTCGGAGGAGAAAGAGCGGATGCCACCAAGGATTCCTGGATGACCTCTTTCCCTAAT 711
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2748 GAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG 2807
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 712 GAGGAGGAAGAGG 724
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2808 GAGGGGCAGGAGG 2820
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-08-194-087-15
; Sequence 15, Application US/08194087
; Patent No. 5879910
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-194-087-15

Query Match          3.2%; Score 77.8; DB 3; Length 10596;
Best Local Similarity 47.6%; Pred. No. 1.8e-10;
Matches 263; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

QY 175 GGCCTGGTGTGGGGCAGCGGGATGAGCCTGCAGGCCACCGCTCAGCCCAAGAGGAGATC 234
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2268 GGGCAGGACGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2327

QY 235 CTGGGGAGCACACGGCTGGTCAGCCAAAGGGCTAGAGGCCCTACGCAGTGAACACAGGCC 294
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2328 GAGGGGCAGGACGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGG 2387

QY 295 GTGCTGCAAAAGCCTGTCCCAGACCATTGAGTGTCTGCAGCAGGGAGGCCATGAGGAAGGG 354
    | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2388 CAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAG 2447

QY 355 CTGGTGCATGAGAAGGCCCGGAGCTTCGCCGTTCTATGGAACATTGAGTCGGGCTG 414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2448 GGGCAGGAGGGGCAGGAGGGGCAGGACGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2507

QY 415 AGTGAGGCCACAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACAGTGGAGTCGGAGAAA 474
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2508 GAGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGG 2567

QY 475 CAGACGTGCGGGCTCAGGTGCGGGCGGCTATGCCAGGAGAACCAGTGGC---TCCGGGAT 531
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2568 CAGGACGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGAG 2627

QY 532 GAGCTGGCTGGCACCCAGCAGCGGGCTACAGCGGCAGTGAACAGGCTGTGGCTCAGCTGGAG 591
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2628 GAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG 2687

QY 592 GAGGAAAAAGACACCTGGAGTTCCTGGGGCAGCTCGGGCAGTATGATGAGGATGGACAT 651
```





```

; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/NSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
;
US-08-728-323A-1

Query Match
Best Local Similarity 3.0%; Score 72.8; DB 4; Length 3489;
Matches 275; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

QY 282 TGAACACCAGGCCGTGCTGCAAAAGCCCTGTCCAGACCATTTAGTGTCTGCAGCAGGGAGG 341
Db 1974 TGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 2033

QY 342 CCATGAGGAAGGCTGTCATGAGAAGSCCGGCAGCTTCGCCGTTCTATGGAAACAT 401
Db 2034 TGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCA 2093

QY 402 TGAGCTCGGGCTGAGTGAGGCCCGCAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACAGT 461
Db 2094 TGAGCAGGAGCAGCAGGATGAGCAGGAGCAGCAGCAGGATGAGCAGGAGCAGGATGAGCA 2153

QY 462 GGAGTCGGAGAAACAGAAAGCTGCGGGCTCAGGTGCGGGCTATGCCAGGAGAACAGTG 521
Db 2154 GCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCA 2213

QY 522 GCTCGGGATGAGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGC 581
Db 2214 GGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAACAGGAGCAGCAGGAGGA 2273

QY 582 TCAGCTGGAGGAGGAGAAAGACCTTGAGTTCTCTGGGCGAGCTGCGGCAGTATGATGA 641
Db 2274 GCAGGAGCAGCAGGAGGAGCAGGAGCAGGAGTATAGAGGAGCAGGAGCAGGAGTTAGAGGA 2333

QY 642 GGATGGACATACCTCGGAGGAGAGAAAGAGCGCGATGCCACCAAGGATTCCCTGGATGACCT 701
Db 2334 TCAGGAGCAGGAGTATAGAGGAGCAGGAGCAGGAGGATTAAGAGGAGCAGGAGGATTAAGA 2393

QY 702 CTTTCTAATGAGGAGGAGAGGAGGACCCAGCAATGGCTTGTCCCGTGGTCAAGGTGCTAC 761
Db 2394 GGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGGAGGAGTT 2453

QY 762 AGCAGCTCAGCAGGGTGGATATGAGATCCCGAGCAAGGTTGGGACGTTGCACAACCTGGT 821
Db 2454 AGAGGAGCAGCAGGAGGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGGAGGA 2513

QY 822 GATCCAGTACCGAGCCCAAGGTCGCTATGAGGTGGCCGTGCCACTCTGTAAAGCAGGCACT 881
Db 2514 GTTAGAGGAGCAGGAGGTGGAAGAGCAGGAGCAGGAGGTGGAAGAGCAGGAGCAGGAGCA 2573

QY 882 AGAGGACCTGGA 893
Db 2574 GGAAGAGCAGGA 2585

RESULT 9
```

```

US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-770-379-20
```

```

Query Match
Best Local Similarity 3.0%; Score 72.8; DB 3; Length 32207;
Matches 275; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

QY 282 TGAACACCAGGCCGTGCTGCAAAAGCCCTGTCCAGACCATTTAGTGTCTGCAGCAGGGAGG 341
Db 20023 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCAGGA 19964

QY 342 CCATGAGGAAGGCTGGTGCATGAGAAGGCCCGGCAGCTTCGCCGTTCTATGGAAACAT 401
Db 19963 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCAGGA 19904

QY 402 TGAGCTCGGGCTGAGTGAGGCCCGCAGGTGATGCTGGCTCTAGCCAGCCACCTGAGCACAGT 461
Db 19903 TGAGCAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCA 19844

QY 462 GGAGTCGGAGAAACAGAAAGCTGCGSGCTCAGGTGCGGGCTATGCCAGGAGAACCAAGTG 521
Db 19843 GCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCA 19784

QY 522 GCTCGGGATGAGCTGGCTGGCAGCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGC 581
Db 19783 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAACAGGAGCAGCAGGAGGA 19724

QY 582 TCAGCTGGAGGAGGAGAAAGACCTGGAGTTCTCTGGGCGAGCTGCGGCAGTATGATGA 641
Db 19723 GCAGGAGCAGCAGGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGGATGAGGA 19664
```

[illegible]

RESULT 11  
 US-08-056-200-93  
 ; Sequence 93, Application US/08056200  
 ; Patent No. 5616500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steiner, Peter M.  
 ; APPLICANT: Lee, Seung-Chul  
 ; APPLICANT: Kim, In-Gyu  
 ; APPLICANT: Chung, Soo-Il  
 ; APPLICANT: Park, Sang-Chul  
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
 ; TITLE OF INVENTION: Methods of Using Same  
 ; NUMBER OF SEQUENCES: 117  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 620 Newport Center Drive, Sixteenth Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/056,200  
 ; FILING DATE: 30-APR-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fedrick, Michael F.  
 ; REGISTRATION NUMBER: 36,799  
 ; REFERENCE/DOCKET NUMBER: NIH054.001A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (714) 760-0404  
 ; TELEFAX: (714) 760-9502  
 ; INFORMATION FOR SEQ ID NO: 93:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9551 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear

Query Match	2.4%	Score	59.6;	DB 1;	Length	7218;
Best Local Similarity	3.0%	pred. No.	6.7e-06;			

MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-056-200-93

Query Match 2.3%; Score 55.6; DB 1; Length 9551;  
Best Local Similarity 46.4%; Pred. No. 7.9e-05;  
Matches 181; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 281 GTGAACACACAGCGCGTGTGCTGCAAAAGCCTGTCCACAGCATTGAGTGTCTGCAGCAGGGAG 340  
Db 3356 GCGAGCAGCAGAGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3415  
QY 341 GCCATGAGGAAGGGCTGGTGCATGAGAAAGGCCCGGAGCTTCCGCGTCTATGGAAACA 400  
Db 3416 AGCAGCAGCTGAGGCGCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3475  
QY 401 TTGAGCTCGGGCTGAGTGAGGCCCGCAGGTGATGCTGCTAGCCAGCCACCTGAGCACAG 460  
Db 3476 AGGAGGAGGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGGAGGAGGAGGAGGAGGAG 3535  
QY 461 TGGAGTCGAGAAACAGAGCTGCGGGCTCAGGTGCGGCGCTATGCCAGGAGAACAGT 520  
Db 3536 AGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAG 3595  
QY 521 GGCTCGGGATGAGTGGTGGCAGCCAGCGCTACAGCGCAGTGAACAGGCTGTGG 580  
Db 3596 TGAGGCGCAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGC 3655  
QY 581 CTCAGCTGAGGAGGAGAAAGAGCAGCCTGGAGTTCTTGGGCGAGCTGCGGCGAGTATGATG 640  
Db 3656 GCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAG 3715  
QY 641 AGGATGGACATACCTCGGAGGAGAAAGAAG 670  
Db 3716 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3745

RESULT 12  
US-08-800-644-93  
Sequence 93, Application US/08800644  
Patent No. 5958752  
GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,644  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/056,200  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-800-644-93

Query Match 2.3%; Score 55.6; DB 4; Length 9551;  
Best Local Similarity 46.4%; Pred. No. 7.9e-05;  
Matches 181; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 281 GTGAACACACAGCGCGTGTGCTGCAAAAGCCTGTCCACAGCATTGAGTGTCTGCAGCAGGGAG 340  
Db 3356 GCGAGCAGCAGAGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3415  
QY 341 GCCATGAGGAAGGGCTGGTGCATGAGAAAGGCCCGGAGCTTCCGCGTCTATGGAAACA 400  
Db 3416 AGCAGCAGCTGAGGCGCGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3475  
QY 401 TTGAGCTCGGGCTGAGTGAGGCCCGCAGGTGATGCTGCTAGCCAGCCACCTGAGCACAG 460  
Db 3476 AGGAGGAGGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGGAGGAGGAGGAGGAGGAG 3535  
QY 461 TGGAGTCGAGAAACAGAGCTGCGGGCTCAGGTGCGGCGCTATGCCAGGAGAACAGT 520  
Db 3536 AGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAG 3595  
QY 521 GGCTCGGGATGAGTGGTGGCAGCCAGCGCTACAGCGCAGTGAACAGGCTGTGG 580  
Db 3596 TGAGGCGCAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGC 3655  
QY 581 CTCAGCTGAGGAGGAGAAAGAGCAGCCTGGAGTTCTTGGGCGAGCTGCGGCGAGTATGATG 640  
Db 3656 GCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAG 3715  
QY 641 AGGATGGACATACCTCGGAGGAGAAAGAAG 670  
Db 3716 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3745

RESULT 13  
US-08-469-802B-3  
Sequence 3, Application US/08469802B  
Patent No. 5741645  
GENERAL INFORMATION:





STATE: MN  
COUNTRY: USA  
ZIP: 55401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,802B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muetling, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 110.00030101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1225  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 195 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-469-802B-2

Query Match 2.0%; Score 48; DB 2; Length 195;  
Best Local Similarity 55.4%; Pred. No. 0.0015;  
Matches 93; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
QY 468 GGAGAAACAGAGCTGCGGGCTCAGGTGCGGGGCTATGCCAGGAGAACCACTGGCTGCG 527  
Db 3 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 62  
QY 528 GGATGAGCTGGCTGGCACCCAGCAGCGGCTACAGCGCAGTGAACAGGCTGTGGCTCAGCT 587  
Db 63 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 122  
QY 588 GGAGGAGGAAAGAACCTGGAGTTCCTGGGGCAGCTGCGGCAGTA 635  
Db 123 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 170

Search completed: August 15, 2000, 04:31:12  
Job time: 13830 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2000, 00:19:33 ; Search time 875.55 Seconds  
(without alignments)  
12354.872 Million cell updates/sec

Title: US-09-036-614A-2  
Perfect score: 2453  
Sequence: 1 GTGAAGTGTGAAGAAGGG.....AGAATCAGTCTTNTNTNNG 2453

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	em_est1:*
2:	em_est2:*
3:	em_est3:*
4:	em_est4:*
5:	em_est5:*
6:	em_est6:*
7:	em_est7:*
8:	em_est8:*
9:	em_est9:*
10:	em_est10:*
11:	em_est11:*
12:	em_est12:*
13:	em_est13:*
14:	em_est14:*
15:	em_est15:*
16:	em_est16:*
17:	em_est17:*
18:	em_est18:*
19:	em_est19:*
20:	gb_est1:*
21:	gb_est2:*
22:	gb_est3:*
23:	gb_est4:*
24:	gb_est5:*
25:	gb_est6:*
26:	gb_est7:*
27:	gb_est8:*
28:	gb_est9:*
29:	gb_est10:*
30:	gb_est11:*
31:	gb_est12:*
32:	gb_est13:*
33:	gb_est14:*
34:	gb_est15:*
35:	gb_est16:*
36:	gb_est17:*
37:	gb_est18:*
38:	gb_est19:*
39:	gb_est20:*
40:	gb_est21:*
41:	gb_est22:*
42:	gb_est23:*
43:	gb_est24:*

44:	gb_est25:*
45:	gb_est26:*
46:	gb_est27:*
47:	gb_est28:*
48:	gb_est29:*
49:	gb_est30:*
50:	gb_est31:*
51:	gb_est32:*
52:	em_est20:*
53:	em_est21:*
54:	em_est22:*
55:	em_est23:*
56:	em_est24:*
57:	em_est25:*
58:	em_est26:*
59:	gb_est33:*
60:	gb_est34:*
61:	gb_est35:*
62:	gb_est36:*
63:	gb_est37:*
64:	gb_est38:*
65:	em_est27:*
66:	em_est28:*
67:	em_est29:*
68:	em_est30:*
69:	gb_est39:*
70:	gb_est40:*
71:	gb_est41:*
72:	gb_est42:*
73:	gb_est43:*
74:	gb_est44:*
75:	em_est31:*
76:	em_est32:*
77:	em_est33:*
78:	em_est34:*
79:	gb_est45:*
80:	gb_est46:*
81:	gb_est47:*
82:	em_est35:*
83:	em_est36:*
84:	em_est37:*
85:	gb_est48:*
86:	gb_est49:*
87:	gb_est50:*
88:	gb_est51:*
89:	gb_est52:*
90:	gb_est53:*
91:	gb_est54:*
92:	gb_est55:*
93:	gb_gss1:*
94:	gb_gss2:*
95:	gb_gss3:*
96:	gb_gss4:*
97:	em_gss1:*
98:	em_gss2:*
99:	em_gss3:*
100:	em_gss4:*
101:	gb_gss5:*
102:	gb_gss6:*
103:	gb_gss7:*
104:	gb_gss8:*
105:	gb_gss9:*
106:	em_gss5:*
107:	em_gss6:*
108:	em_gss7:*
109:	em_gss8:*
110:	em_gss9:*
111:	em_gss10:*
112:	em_gss11:*
113:	gb_gss10:*
114:	gb_gss11:*
115:	em_gss12:*
116:	gb_gss12:*





||||| 434 ATCAACCGCGCCCGAGGTCTGNTGGGT-CCCCACCCACAGCCCTCACAGCATTC 376

QY 2090 CCATTGCTCTGGCTCTTCCACCCCTAGTGGGACAGTGAAGGGGAGCAGTTTAACCA 2149

Db 375 CCATTGCTCTGGCTCTTCCACCCCTAGTGGGACAGTGAAGGGGAGCAGTTTAACCA 316

QY 2150 GAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTCTTAGGAAGGA 2209

Db 315 GAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTCTTAGGAAGGA 256

QY 2210 CCCTCAGGACACCCCTCTCTGCACCCTGTGGTCTCTAGAGTAGCTAGCTCTGAGGCCCA 2269

Db 255 CCCTCAGGACACCCCTCTCTGCACCCTGTGGTCTCTAGAGTAGCTAGCTCTGAGGCCCA 196

QY 2270 AGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCTGCTGCTGGCTTTTCACTCA 2329

Db 195 AGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCTGCTGCTGGCTTTTCACTCA 136

QY 2330 GAGGGTTGGGGGCTGGCCAGCAAGCTGCCTTGCCCTGCGGCTCTTACTCCCTCCCTCT 2389

Db 135 GAGGGTTGGGGGCTGGCCAGCAAGCTGCCTTGCCCTGCGGCTCTTACTCCCTCCCTCT 76

QY 2390 GCTGTCTCACTTCAGGTCCATGTATTTCACCTTTTCTTAATAAAGAATCAG 2441

Db 75 GCTGTCTCACTTCAGGTCCATGTATTTCACCTTTTCTTAATAAAGAATCAG 24

RESULT 2

AW371381 500 bp mRNA EST 04-FEB-2000

LOCUS RC0-BT0291-011199-011-e03 BT0291 Homo sapiens cDNA, mRNA sequence.

DEFINITION AW371381

ACCESSION AW371381

VERSION AW371381.1 GI:6876035

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 500)

TITLE HCGP <http://www.ludwig.org.br/ORESTES>.

JOURNAL The FAPESP/LICR Human Cancer Genome Project

COMMENT Unpublished (1999)

On Apr 30, 1999 this sequence version replaced gi:4728180.

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC0&t2=RC0-BT0291-011199-011-e03&t3=1999-11-01&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 10

High quality sequence stop: 70.

Location/Qualifiers

1..500

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BT0291"

/dev\_stage="Adult"

/note="Organ: breast; vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 139 a 123 c 153 g 85 t

ORIGIN

Query Match 19.6%; Score 480; DB 71; Length 500;

Best Local Similarity 98.8%; Pred. No. 6.4e-108;

Matches 494; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1134 GGAGATTCGAGAAAAGTCTCTGGGCACGAATCATCCAGATGTGGCAAAACAGCTGAACAA 1193

Db 1 GGAGATTCGAGAAAAGTCTCTGGGCATGAATCATCCAGATGTGGCAAAACAGCTGAACAA 60

QY 1194 CCTGGCCCTCTTGTGCCAAAACAGGGCAAGTATGAGGCCGTGGAACGCTACTACCAAGC 1253

Db 61 CCTGGCCCTCTTGTGCCAAAACAGGGCAAGTATGAGGCCGTGGAACGCTACTACCAAGC 120

QY 1254 AGCACTGGCCATCTACGAGGGGCAGCTGGGGCCGACAAACCC-TAATGTAGCCCCGACCA 1312

Db 121 AGCACTGGCCATCTACGAGGGGCAGCTGGTCCCGACACCCCTTAATGTAGCCCCGACCA 180

QY 1313 AGAACAACTGGCTTCCTGTCTTACCTGAAACAGGGCAATATGCTGAGGCTGAGACACTAT 1372

Db 181 AGAACAACTGGCTTCCTGTCTTACCTGAAACAGGGCAATATGCTGAGGCTGAGACACTAT 240

QY 1373 ACAAGAGATCCTGACCCGTCGCCATGTACAGGAGTTTGGGTCTGTGGATGATGACCAACA 1432

Db 241 ACAGAGAGATCCTGACCCGTCGCCATGTACAGGAGTTTGGGTCTGTGGATGATGACCAACA 300

QY 1433 AGCCCATCTGGATGCATCAGAGGAGCGGGAGGAAATGAGCAAAAAGCCGGCACCATGAGG 1492

Db 301 AGCCCATCTGGATGCATCAGAGGAGCGGGAGGAAATGAGCAAAAAGCCGGCACCATGAGG 360

QY 1493 GTGGGACACCCCTATGCTGAGTATGGAGGCTGGTACAGGCCCTGCAAAGTGAGCAGCCCCA 1552

Db 361 GTGGGACACCCCTATGCTGAGTATGGAGGCTGGTACAGGCCCTGCAAAGTGAGCAGCCCCA 420

QY 1553 CAGTGAACACTACTCTGAGAAACCTGGGAGCTCTGTATAGGCCGCCAGGAAAGCTGGAGG 1612

Db 421 CAGTGAACACTACTCTGGGAAACCTGGGAGCTCTGTATAGGCCGCCAGGAAAGCTGGAGG 480

QY 1613 CTGCTGAGACCCCTGGAGGAA 1632

Db 481 CTGCTGAGACCCCTGGAGGAA 500

RESULT 3

AI041608/c

LOCUS ox92e08.x1 Soares senescent\_fibroblasts\_NbHSF Homo sapiens cDNA

DEFINITION clone IMAGE:1663814 3', mRNA sequence.

ACCESSION AI041608

VERSION AI041608.1 GI:3280802

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 463)

TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3187022.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: [Robert\\_Strausberg@nih.gov](mailto:Robert_Strausberg@nih.gov)

This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Insert Length: 972 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 419.

Location/Qualifiers

1..463

/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone="IMAGE:1663814"  
/clone\_lib="Soares\_senescent\_fibroblasts\_NbHSP"  
/tissue\_type="senescent fibroblast"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT7T3D (Pharmacia) with a modified  
polylinker V\_TYPE: phagemid; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5',  
TGTACCAATCTGAAGTGGGAGCGCGCATTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo."  
BASE COUNT 106 a 107 c 154 g 96 t  
ORIGIN

Query Match 18.4%; Score 451; DB 34; Length 463;  
Best Local Similarity 100.0%; Pred. No. 8.8e-101;  
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 TCAGTGCCAGCACCATGGACCTCTCTTCAAGCAGCTGACATTC AACCCGCGCCCGCAGGTC 2050  
Db 463 TCAGTGCCAGCACCATGGACCTCTCTTCAAGCAGCTGACATTC AACCCGCGCCCGCAGGTC 404  
  
QY 2051 TGCTGGGTCCCCCACCCACACAGCCCTCACAGCATTCCTCCATTGCTCGGCTCTTCCC 2110  
Db 403 TGCTGGGTCCCCCACCCACACAGCCCTCACAGCATTCCTCCATTGCTCGGCTCTTCCC 344  
  
QY 2111 CACCCCTAGTGGGACAGTGAAGGGGAGCAGTTTAACCAGAAGATTGCTGCTGCCCTTAG 2170  
Db 343 CACCCCTAGTGGGACAGTGAAGGGGAGCAGTTTAACCAGAAGATTGCTGCTGCCCTTAG 284  
  
QY 2171 GGTCCTCAGCTCCCTCCTCAGGAATCCCTCTTAGAAGGACCCCTCAGGACACCCCTCTCTGC 2230  
Db 283 GGTCCTCAGCTCCCTCCTCAGGAATCCCTCTTAGAAGGACCCCTCAGGACACCCCTCTCTGC 224  
  
QY 2231 ACCCTGTGGTCCTCTAGACTAGCTAGCTCTGAGGCCCAAGGTGGGTACAAAGCAGGTAT 2290  
Db 223 ACCCTGTGGTCCTCTAGACTAGCTAGCTCTGAGGCCCAAGGTGGGTACAAAGCAGGTAT 164  
  
QY 2291 GGCCCTCAGAGATGCAGCTGCTGCTGGCTTTTCAGTCAGAGGGTTGGGGCTGGCCAGC 2350  
Db 163 GGCCCTCAGAGATGCAGCTGCTGCTGGCTTTTCAGTCAGAGGGTTGGGGCTGGCCAGC 104  
  
QY 2351 CAAGCTGCCTTGCCCTGGCGCTCTTACTCCCTCCTCTGCTGCTCAGTCAGGTCCAT 2410  
Db 103 CAAGCTGCCTTGCCCTGGCGCTCTTACTCCCTCCTCTGCTGCTCAGTCAGGTCCAT 44  
  
QY 2411 GTATTTCACCTTTTCTTAATAAAGAATCAG 2441  
Db 43 GTATTTCACCTTTTCTTAATAAAGAATCAG 13

RESULT 4  
AI814907/c  
LOCUS  
DEFINITION wk68e02.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2420570 3',  
mRNA sequence. EST 07-MAR-2000  
AI814907  
ACCESSION AI814907  
VERSION AI814907.1 GI:5426122  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 747)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT On Jun 15, 1998 this sequence version replaced gi:3226427.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Life Technologies catalog #: 11548-013  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 899 Std Error: 0.00  
Seq primer: -400p from Gibco  
High quality sequence stop: 390.  
FEATURES  
Location/Qualifiers  
1..747  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2420570"  
/clone\_lib="NCI\_CGAP\_Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.72 kb. Life Technologies catalog #:  
11548-013"

BASE COUNT 159 a 164 c 233 g 167 t 24 others  
ORIGIN

Query Match 17.7%; Score 434.4; DB 44; Length 747;  
Best Local Similarity 93.7%; Pred. No. 1.2e-96;  
Matches 503; Conservative 0; Mismatches 28; Indels 6; Gaps 5;

QY 1905 GCCTCCAGCAGCAACATGAAGCGAGCGCTCCTTGAACATATCTGAACCAACCTAGTGC 1964  
Db 540 GCCTCCAGCAGCAACAT-AAGCGAGCAGCTTCTTGAACATATCTGAAC--AACTTAGTC 484  
  
QY 1965 AGCACCCCTCCAGGTCTCCGGGGCCCTCAGTGCAGCACCATTGGACCTCTCTCAAGCAG 2024  
Db 483 AGCACCCCTCAAGGTCTCCA-GGGCCTAAGTCCAAGCACAAAT-GACCTCTTCAAGCAG 426  
  
QY 2025 CTGACATTCAACCCCGCCCCCAGGTCTGCTGGTCCCTCCACCCACAGCCCTCACAGC 2084  
Db 425 NTACAATTCAACCCGCGCCNCCAGGTNTGNTGGNT-CCCCACCCACAGCCCTCACAGC 367  
  
QY 2085 ATTCCCATTTGCTCCTGGCTCTTCCCCACCCCTAGGTGGACAGTGAAGGGGAGCAGTTT 2144  
Db 366 ATTCCCATTTGCTCCTGGCTCTTCCCCACCCCTAGGTGGACAGTGAAGGGGAGCAGTTT 307  
  
QY 2145 AACAGAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTTTAGG 2204  
Db 306 AACAGAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTTTAGG 247  
  
QY 2205 AAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGCTGGTCCCTCTAGAGTAGCTCTGAGG 2264  
Db 246 AAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGCTGGTCCCTCTAGAGTAGCTCTGAGG 187  
  
QY 2265 CCCCAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGGCTTTTC 2324  
Db 186 CCCCAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCCTGCTGGCTTTTC 127  
  
QY 2325 AGTCAGAGGGTTGGGGCTGGCCAGCCCAAGCTGCCCTTGCCCTGGCCGCTCTTACTCCCTC 2384  
Db 126 AGTCAGAGGGTTGGGGCTGGCCAGCCCAAGCTGCCCTTGCCCTGGCCGCTCTTACTCCCTC 67  
  
QY 2385 CCTCTGCTGTCTCACTTCAGGTCCATGTATTTCACTTTTCTTAAATAAAGAATCAG 2441  
Db 66 CCTCTGCTGTCTCACTTCAGGTCCATGTATTTCACTTTTCTTAAATAAAGAATCAG 10

RESULT 5  
AI076952/c  
LOCUS AI076952 450 bp mRNA EST 27-AUG-1998

DEFINITION ov20h09.x1 NCI\_CGAP\_Br2 Homo sapiens cDNA clone IMAGE:1637921 3',  
 mRNA sequence.  
 ACCESSION AI076952  
 VERSION AI076952.1 GI:3404781  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 450)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2285573.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Insert length: 614 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham.  
 FEATURES  
 Location/Qualifiers  
 1..450  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1637921"  
 /clone\_lib="NCI\_CGAP\_Br2"  
 /sex="female, pooled"  
 /tissue\_type="breast"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from pooled bulk  
 breast tumor tissue, and was then primed with a Not I -  
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
 RI adaptors (Pharmacia), digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT7T3  
 vector. This library is the normalized version of  
 NCI\_CGAP\_Brl.1. Library was constructed by Bento Soares  
 and M. Fatima Bonaldo."  
 .  
 BASE COUNT 104 a 102 c 147 g 96 t 1 others  
 ORIGIN  
 Query Match 17.4%; Score 428; DB 35; Length 450;  
 Best Local Similarity 99.8%; Pred. No. 4.1e-95;  
 Matches 428; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2013 CTCTTCAAGCAGCTGACATTCAACCCGGCCCCCAGGTCTGTGGTCCCCCACCACCCAC 2072  
 |||  
 Db 450 CTCTTCAAGCAGCTGACATTCAACCCGGCCCCCAGGTCTGTGGTCCCCCACCACCCAC 391  
 QY 2073 AGCCCTCACAGCATCCCCCATTGCTCCTGGCTCTTCCCCACCCCTAGGTGGGACAGTGAA 2132  
 |||  
 Db 390 AGCCCTCACAGCATCCCCCATTGCTCCTGGCTCTTCCCCACCCCTAGGTGGGACAGTGAA 331  
 QY 2133 GGGGAGCAGTTTAAACAGAAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGA 2192  
 |||  
 Db 330 GGGGAGCAGTTTAAACAGAAAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGA 271  
 QY 2193 ATCCCTCTTAGGAAGGACCCTCAGGACACCCCTCTCTGCACCCTGTGGTCTCTAGAGTAG 2252  
 |||  
 Db 270 ATCCCTCTTAGGAAGGACCCTCAGGACACCCCTCTCTGCACCCTGTGGTCTCTAGAGTAG 211  
 QY 2253 CTAGCTCTGAGGCCCCAAGGTGGGTACAAAGCAGGTATGCCCTCAGAGATGCAGCCTGC 2312  
 |||  
 Db 210 CTAGCTCTGAGGCCCCAAGGTGGGTACAAAGCAGGTATGCCCTCAGAGATGCAGCCTGC 151



```
Db 441 TCTCTTAAGCAGCTGACATTCAACCCGGCCCCCAGGTCTGNTGGTCCCCCACCACCCCA 382
QY 2072 CAGCCCTCACAGCATCCCCATTGCTCCGGCTCTTCCCCACCCCTAGGTGGGACAGTGA 2131
Db 381 CAGCCCTCACAGCATCCCCATTGCTCCGGCTCTTCCCCACCCCTAGGTGGGACAGTGA 322
QY 2132 AGGGGAGCAGTTTAACCAAGAAATTGCTGTGCCCTTAGGGTCTCAGCTCCCTCCTCAGG 2191
Db 321 AGGGGAGCAGTTTAACCAAGAAATTGCTGTGCCCTTAGGGTCTCAGCTCCCTCCTCAGG 262
QY 2192 AATCCCTCTTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGGGTCTCTAGAGTA 2251
Db 261 AATCCCTCTTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGGGTCTCTAGAGTA 202
QY 2252 GCTAGCTCTGAGSCCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTG 2311
Db 201 GCTAGCTCTGAGSCCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTG 142
QY 2312 CTGCTGGCTTTTCAGTCAGAGGGTTGGGGCTGGCCAGCCAAAGCTGCCCTGGCCG 2371
Db 141 CTGCTGGCTTTTCAGTCAGAGGGTTGGGGCTGGCCAGCCAAAGCTGCCCTGGCCG 82
QY 2372 CTCTTACTCCCTCCCTCTGCTGTCTCACTCAGGTCCATGTATTTCACTTTTCTTAATA 2431
Db 81 CTCTTACTCCCTCCCTCTGCTGTCTCACTCAGGTCCATGTATTTCACTTTTCTTCAATA 22
QY 2432 AAAGAATCAG 2441
Db 21 AAAGAATCAG 12

RESULT 7
AW072690/c
LOCUS
DEFINITION xa41h04.x1 NCI_CGAP_Sar4 Homo sapiens cDNA clone IMAGE:2569399 3',
mRNA sequence.
ACCESSION AW072690
VERSION AW072690.1 GI:6027688
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 758)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189325.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 389.
Location/Qualifiers
1. .758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2569399"
/clone_lib="NCI_CGAP_Sar4"
/tissue_type="five pooled sarcomas, including myxoid
liposarcoma, solitary fibrous tumor, malignant fibrous
histiocytoma, gastrointestinal stromal tumor, and
```

```
mesothelioma"
/lab_host="DH10B"
/note="Organ: connective tissue; Vector: pCMV-SPORT6;
Site_1: Sali; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Library constructed by Life
Technologies."
BASE COUNT 180 a 157 c 261 g 152 t 8 others
ORIGIN

Query Match 17.4%; Score 426.6; DB 63; Length 758;
Best Local Similarity 94.3%; Pred. No. le-94;
Matches 450; Conservative 0; Mismatches 25; Indels 2; Gaps 1;

QY 1965 AGCACCCCTCCAGGTCTCCCGGGCCCTCAGTCCAGCACCATGGACCTCTCTTCAAGCAG 2024
Db 485 AGCCCCCTTCAGGTCTTC--GTGGCTTCAGTGCAGCCCCCATGGACCTCTCTCNAGCAAG 428
QY 2025 CTGACATTCAACCCGGCCCCCAGGTCTGCTGGGTCCCCCACCCACAGCCCTCACAGC 2084
Db 427 CTGACATTCAACCCGGCCCCCANNGTNTGTTGGTCCCNCCACCCACAGCCCTCACAGC 368
QY 2085 ATTCCCATTTGCTCTGGCTCTTCCACCCCTTAGTGGGACAGTGAAGGGGAGCAGTTT 2144
Db 367 ATTCCCATTTGCTCTGGCTCTTCCACCCCTTAGTGGGACAGTGAAGGGGAGCAGTTT 308
QY 2145 AACCAGAAGATTGCTGCTGCCCTTAGGTCTCAGTCCCTCCTCAGGAATCCCTCTTAGG 2204
Db 307 AACCAGAAGATTGCTGCTGCCCTTAGGTCTCAGTCCCTCCTCAGGAATCCCTCTTAGG 248
QY 2205 AAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGTGTCCTCTAGAGTAGTACTCTGAGG 2264
Db 247 AAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGTGTCCTCTAGAGTAGTACTCTGAGG 188
QY 2265 CCCCAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTGCTGCTGGCTTTTC 2324
Db 187 CCCCAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCAGCCTGCTGCTGGCTTTTC 128
QY 2325 AGTCAGAGGGTTGGGGCTGGCCAGCCAAAGCTGCCTTGGCCCTGCTTACTCCCTC 2384
Db 127 AGTCAGAGGGTTGGGGCTGGCCAGCCAAAGCTGCCTTGGCCCTGCTTACTCCCTC 68
QY 2385 CCTCTGCTCTCATTTCAGGTCCATGATTTTCACTTTTCTTAATAAAAGAATCAG 2441
Db 67 CCTCTGCTCTCATTTCAGGTCCATGATTTTCACTTTTCTTAATAAAAGAATCAG 11

RESULT 8
AA678056/c
LOCUS
DEFINITION AA678056 425 bp mRNA EST 19-DEC-1997
clone IMAGE:431839 3', mRNA sequence.
ACCESSION AA678056
VERSION AA678056.1 GI:2658578
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 425)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through L3NL ; coptact the
```



IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 410.  
Location/Qualifiers  
1. .425  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:431839"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI; This is a subcloned version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATAAGATCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT	99 a	101 c	143 g	82 t
ORIGIN				
Query Match	16.7%;	Score 410;	DB 29;	Length 425;
Best Local Similarity	99.8%;	Pred. No. 1.e-90;		
Matches 421;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;
QY	2020	AGCAGCTGACATTCAACCCGGCCCCAGGTCGTGGTCCCCACCCACAGCCCTC	2079	
Db	425	AGCAGCTGACATTCAACCCGGCCCCAGGTCGTGGTCCCCACCCACAGCCCTC	367	
QY	2080	ACAGCATTCCTCCATTGCTCCTGGCTCTCCACCCCTAGGTGGGACAGTGAAGGGGAGC	2139	
Db	366	ACAGCATTCCTCCATTGCTCCTGGCTCTCCACCCCTAGGTGGGACAGTGAAGGGGAGC	307	
QY	2140	AGTTTAAACAGAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	2199	
Db	306	AGTTTAAACAGAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTC	247	
QY	2200	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGTGGTCCCTCTAGAGTAGCTAGCTC	2259	
Db	246	TTAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGTGGTCCCTCTAGAGTAGCTAGCTC	187	
QY	2260	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCGCCTGCTGCTGGC	2319	
Db	186	TGAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGCGCCTGCTGCTGGC	127	
QY	2320	TTTTTCAGTCAGAGGGTTGGGGCTGGCCAGCAAGCTGCCCTTGCCTGGCCGCTCTTACT	2379	
Db	126	TTTTTCAGTCAGAGGGTTGGGGCTGGCCAGCAAGCTGCCCTTGCCTGGCCGCTCTTACT	67	
QY	2380	CCCTCCCTCTGCTGCTCCTCAGTTCAGGTCCATGTATTTTCTTTTCTTAATAAAGAATC	2439	
Db	66	CCCTCCCTCTGCTGCTCCTCAGTTCAGGTCCATGTATTTTCTTTTCTTAATAAAGAATC	7	
QY	2440	AG 2441		
Db	6	AG 5		

RESULT 9.  
AA639663/c  
LOCUS AA639663 428 bp mRNA EST 23-OCT-1997  
DEFINITION nq83912.s1 NCI\_CGAP\_Co9 Homo sapiens cDNA clone IMAGE:1158982 3', mRNA sequence.  
ACCESSION AA639663  
VERSION AA639663.1 GI:2563442  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 428)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/hcicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2285718.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: -40ml3 fwd. ET from Amersham.  
Location/Qualifiers  
1. .428  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1158982"  
/clone\_lib="NCI\_CGAP\_Co9"  
/tissue\_type="colon tumor RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from RER+ colon tumor, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo (Soares4)."

BASE COUNT	101 a	101 c	145 g	81 t
ORIGIN				
Query Match	16.7%;	Score 409;	DB 29;	Length 428;
Best Local Similarity	99.8%;	Pred. No. 2e-90;		
Matches 420;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;
QY	2021	GCAGCTGACATTCAACCCGGCCCCAGGTCGTGGTCCCCACCCACAGCCCTCA	2080	
Db	428	GCAGCTGACATTCAACCCGGCCCCAGGTCGTGGTCCCCACCCACAGCCCTCA	370	
QY	2081	CAGCATTCCTCCATTGCTCCTGGCTCTCCACCCCTAGGTGGGACAGTGAAGGGGAGCA	2140	
Db	369	CAGCATTCCTCCATTGCTCCTGGCTCTCCACCCCTAGGTGGGACAGTGAAGGGGAGCA	310	
QY	2141	GTTTAAACAGAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTCT	2200	
Db	309	GTTTAAACAGAGATTGCTGCTGCCCTTAGGGTCTCAGCTCCCTCCTCAGGAATCCCTCT	250	
QY	2201	TAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGTGGTCTCTAGAGTAGCTAGCTCT	2260	
Db	249	TAGGAAGGACCCCTCAGGACACCCCTCTCTGCACCCCTGTGGTCTCTAGAGTAGCTAGCTCT	190	
QY	2261	GAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGAGCCCTGCTGGCT	2320	
Db	189	GAGGCCCAAGGTGGGTACAAAGCAGGTATGGCCCTCAGAGATGAGCCCTGCTGGCT	130	
QY	2321	TTTCAGTCAGAGGGTTGGGGCTGGCCAGCAAGCTGCCCTGCGCCGCTCTTACTC	2380	
Db	129	TTTCAGTCAGAGGGTTGGGGCTGGCCAGCAAGCTGCCCTGCGCCGCTCTTACTC	70	
QY	2381	CCTCCCTCTGCTGCTCCTCAGTTCAGGTCCATGTATTTTCTTTTCTTAATAAAGAATCA	2440	
Db	69	CCTCCCTCTGCTGCTCCTCAGTTCAGGTCCATGTATTTTCTTTTCTTAATAAAGAATCA	10	













